

密级: _____



中国科学院大学
University of Chinese Academy of Sciences

博士学位论文

基于深度学习的医学图像内容理解关键技术研究

作者姓名: _____ 陶攀

指导教师: _____ 付忠良 研究员

_____ 中国科学院成都计算机应用研究所

学位类别: _____ 工学博士

学科专业: _____ 计算机软件与理论

培养单位: _____ 中国科学院成都计算机应用研究所

2017 年 12 月

Research on Key Technologies in Medical Image Processing

Based on Deep Learning

by
Pan Tao

A thesis submitted to
The University of Chinese Academy of Sciences
in partial fulfillment of the requirements
for the degree of
PhD of Computer Software and Theory

Chengdu Institute of Computer Applications
Chinese Academy of Sciences

December, 2017

学位论文独创性声明

本人郑重声明：我所呈交的学位论文是本人在导师指导下进行的研究工作及所取得的研究成果。尽我所知，除了文中已经标注引用的内容外，本论文中不含其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中作了明确的说明或致谢。本人知道本声明的法律结果由自己承担。

作者签名：_____ 日期：_____

关于学位论文使用授权的说明

本人完全了解中国科学院成都计算机应用研究所有关保留、使用学位论文的规定，即：中国科学院成都计算机应用研究所有权保留送交论文的复印件，允许论文被查阅和借阅；可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。

(保密的论文在解密后应遵守此规定)

作者签名：_____ 导师签名：_____ 日期：_____

摘 要

关键词：中国科学院大学，学位论文，L^AT_EX 模板

Abstract

This paper is a help documentation for the L^AT_EX class ucasthesis, which is a thesis template for the University of Chinese Academy of Sciences. The main content is about how to use the ucasthesis, as well as how to write thesis efficiently by using L^AT_EX.

Keywords: University of Chinese Academy of Sciences (UCAS), Thesis, L^AT_EX Template

目 录

摘要	vii
Abstract	ix
目录	xi
图形列表	xiii
表格列表	xv
符号列表	xvii
第一章 绪论	1
1.1 研究背景及现实意义	1
1.2 国内外研究现状及难点	4
1.3 全文结构及创新点	6
1.4 本章小结	6
第二章 医学影像与深度学习相关知识	7
2.1 医学影像	7
2.2 人工神经网络	7
2.3 卷积神经网络	7
2.4 深度卷积神经网络	8
2.5 本章小结	8
第三章 超声心动图切面的识别方法	9
3.1 引言	9
3.2 Deep-Echo 模型	10
3.3 实验结果和分析	13
3.4 本章小结	16
第四章 空间金字塔分解的深度可视化方法	17
4.1 引言	17
4.2 可视化方法的数学模型	18
4.3 梯度更新的可视化方法	19

4.4 空间金字塔分解	20
4.5 实验结果分析和讨论	22
4.6 本章小结	26
第五章 医学计算机辅助检测方法	27
5.1 引言	27
5.2 区域卷积神经网络概览	28
5.3 候选区域生成网络及其改进	30
5.4 实验结果分析和讨论	34
5.5 本章小结	36
第六章 医学图像的分割方法	37
6.1 引言	37
6.2 初始位置定位和特征点标注	37
6.3 AAM 模型和 CLM 模型	37
6.4 结合卷积网络特征的形状对齐模型	37
6.5 实验结果分析和讨论	37
6.6 本章小结	37
第七章 总结与展望	39
参考文献	41
攻读学位期间发表的学术论文与科研成果	47
致 谢	51

图形列表

3.1	Deep-Echo 模型结构示意图	11
3.2	七类标准切面超声心动图及数量分布	13
3.3	不同数据量的平均分类精度	14
3.4	各类切面的原图和显著性热力图	15
3.5	深度模型泛化性能可视化分析	16
4.1	高斯和拉普拉斯金字塔	21
4.2	不同模型类别可视化实验结果	23
4.3	不同正则化方法的可视化效果	24
4.4	金字塔分解正则化可视化效果	25
4.5	引入类别激活图的可视化	25
5.1	左上：引入空间不变性的 anchor 机制 左下：空间变换网络右：Faster R-CNN 带仿射变换的检测模型框架	30
5.2	考虑物体朝向的区域生成网络模型结构示意图，图中 conv 表示卷积层，pool 表示池化层，右侧的带仿射变换的区域生成网络模型同最终的检测模型一致，其中的 FC 表示全连接层，softmax class loss 表示多任务损失中的分类损失，Bbox.reg loss 表示候选框回归定位损失，Rotation loss 表示文中的针对变换参数 的 Von Mise 损失。	31

表格列表

3.1 不同模型分类精度比较	15
5.1 算法 1 物体检测	29

符号列表

Characters

Symbol	Description	Unit
R	the gas constant	$\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$
C_v	specific heat capacity at constant volume	$\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$
C_p	specific heat capacity at constant pressure	$\text{m}^2 \cdot \text{s}^{-2} \cdot \text{K}^{-1}$
E	specific total energy	$\text{m}^2 \cdot \text{s}^{-2}$
e	specific internal energy	$\text{m}^2 \cdot \text{s}^{-2}$
h_T	specific total enthalpy	$\text{m}^2 \cdot \text{s}^{-2}$
h	specific enthalpy	$\text{m}^2 \cdot \text{s}^{-2}$
k	thermal conductivity	$\text{kg} \cdot \text{m} \cdot \text{s}^{-3} \cdot \text{K}^{-1}$
T	temperature	K
t	time	s
p	thermodynamic pressure	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
\hat{p}	hydrostatic pressure	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
\mathbf{f}_b	body force	$\text{kg} \cdot \text{m}^{-2} \cdot \text{s}^{-2}$
S	boundary surface	m^2
V	volume	m^3
\mathbf{V}	velocity vector	$\text{m} \cdot \text{s}^{-1}$
u	x component of velocity	$\text{m} \cdot \text{s}^{-1}$
v	y component of velocity	$\text{m} \cdot \text{s}^{-1}$
w	z component of velocity	$\text{m} \cdot \text{s}^{-1}$
c	speed of sound	$\text{m} \cdot \text{s}^{-1}$
\mathbf{r}	position vector	m
\mathbf{n}	unit normal vector	1
$\hat{\mathbf{t}}$	unit tangent vector	1
$\tilde{\mathbf{t}}$	unit bitangent vector	1
C_R	coefficient of restitution	1
Re	Reynolds number	1
Pr	Prandtl number	1

Ma	Mach number	1
α	thermal diffusivity	$\text{m}^2 \cdot \text{s}^{-1}$
μ	dynamic viscosity	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-1}$
ν	kinematic viscosity	$\text{m}^2 \cdot \text{s}^{-1}$
γ	heat capacity ratio	1
ρ	density	$\text{kg} \cdot \text{m}^{-3}$
σ_{ij}	stress tensor	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
S_{ij}	deviatoric stress tensor	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
τ_{ij}	viscous stress tensor	$\text{kg} \cdot \text{m}^{-1} \cdot \text{s}^{-2}$
δ_{ij}	Kronecker tensor	1
I_{ij}	identity tensor	1

Operators

Symbol	Description
Δ	difference
∇	gradient operator
δ^\pm	upwind-biased interpolation scheme

Abbreviations

Acronym	Description
ANFO	Ammonium Nitrate Fuel Oil
CFD	Computational Fluid Dynamics
CFL	Courant-Friedrichs-Lowy
CJ	Chapman-Jouguet
EOS	Equation of State
JWL	Jones-Wilkins-Lee
TVD	Total Variation Diminishing
WENO	Weighted Essentially Non-oscillatory
ZND	Zel'dovich-von Neumann-Doering

第一章 絮论

1.1 研究背景及现实意义

1.1.1 医学图像内容理解的研究背景

医学影像技术已经彻底变革了医疗卫生系统，成熟的成像模式不断完善，新技术不断涌现，能更早更有效地诊断身心健康状况，为治疗计划提供信息。用于医疗诊断的成像使医生能够更早地发现疾病并改善患者预后，介入或术中成像有助于消除和治愈许多检测到的疾病。但世界医疗卫生系统每天都会浪费大量的资源和时间，对医学图像内容的错误理解会造成错误诊断，导致很多不必要的额外检查，导致治疗计划的延迟，大大减少了如果早期正确发现的生存或缓解率。

机器学习已被用于医学成像，一些医疗保健和技术创新者正在通过实验人工智能（AI）和机器学习来协作并试图改变我们目前的现实。计算机及其运行的算法可以比人类科学家或医学专业人员更快，更准确地提取大量数据，挖掘模式和预测，加强疾病诊断，提供治疗计划，加强公共健康和安全。

机器现在正在学习如何读取 CT 扫描和其他影像诊断测试来识别异常。虽然有人预测放射科医生的结局，但是也有人认为 AI 是放射科医生的助手。图像中解剖结构的准确分类和定位是全自动的基于图像的胎盘异常诊断的前兆。对于通常在临床筛查和风险评估诊所获得的胎盘超声图像，这些结构可能具有相当模糊的界限和低对比度，并且即使对于有经验的临床医生来说，图像级解释也是具有挑战性和耗时的任务。机器学习是识别可应用于医学图像的图案的技术。虽然这是一个强大的工具，可以帮助提供医疗诊断，但可能会被误用。机器学习通常始于机器学习算法系统，该系统计算被认为在进行感兴趣的预测或诊断中是重要的图像特征。然后，机器学习算法系统识别这些图像特征的最佳组合，以对图像进行分类或计算给定图像区域的一些度量。有几种方法可以使用，每种方法都有不同的优缺点。大多数这些机器学习方法都有开源版本，使得它们易于尝试应用于图像。测量算法性能的几个指标存在；但是，必须意识到可能导致误导性指标的可能的相关缺陷。最近，深度学习已经开始被使用。这种方法具有不需要图像特征识别和计算作为第一步的优点；相反，功能被识别为学习过程的一部分。

1.1.2 课题研究意义

近来，人工智能（也称为深度学习，机器学习或人工神经网络）将如何帮助临床医生的第一个具体例子现在正在商业化。机器学习软件将作为一个非常有经验的临床助理，使医生工作流程更有效率。这些系统可能会为临床医生的工作方式带

来范例转变，努力显着提高工作流程效率，同时提高护理和患者的吞吐量。今天，医生和临床医生面临的最大问题之一就是过多的病人信息过多。电子数据的快速积累归功于电子医疗记录（EMR）的出现以及捕获了以前没有记录的或者至少不容易被数据挖掘的关于病人的各种数据。这包括成像数据，检查和程序报告，实验室值，病理报告，波形，从植入式电生理设备自动下载的数据，从成像和诊断系统本身传输的数据，以及 EMR 中输入的信息，入院，出院和转移（ADT），医院信息系统（HIS）和计费软件。在接下来的几年里，使用双向病人门户网站将会有更多的数据爆发，病人可以将他们自己的数据和图像上传到他们的 EMR 中。这将包括与他们的手机拍摄的伤口部位愈合的图像，以减少需要现场后续办公室访问。它还将包括药物依从性跟踪，血压和体重日志，血糖，抗凝剂 INR 和其他家庭监测测试结果，以及来自应用程序，可穿戴设备和不断发展的物联网（IoT）的活动跟踪，以帮助保持患者健康。医生们把所有这些数据都比作饮用水，因为它是压倒性的。许多人认为通过大量数据来挑选临床相关或可操作性是非常困难或不可能的。事情很容易通过裂缝掉下来，或者由于病人随访而丢失。如果增加诸如增加患者数量，降低报销额，捆绑支付以及从服务费转换为按价值计费的报销系统等因素，这个问题就更加复杂化了。这是人工智能在未来几年将发挥关键作用的地方。人工智能不会诊断患者，也不会取代医生 - 这将增加他们找到需要照顾患者的关键数据的能力，并以简洁易懂的格式呈现患者。在二月的 2017 年健康信息和管理系统协会（HIMSS）年度会议上，几家供应商展示了这种类型的 AI 如何工作的一些具体例子。IBM / Merge，飞利浦，爱克发和西门子已经开始将 AI 集成到他们的医疗成像软件系统中。通用电气公司使用人工智能的元素来显示预测分析软件，以便在有人呼叫病人或病人数量增加时，对影像科室产生影响。Vital 展示了一个用于成像设备利用率的类似工作中预测分析软件。包括几家分析公司和创业公司在内的其他公司则展示了使用 AI 快速筛选大量大数据的软件，或者为适当的使用标准提供即时的临床决策支持，最好的测试或成像来进行诊断甚至提供差异诊断。飞利浦将 AI 作为其具有自适应智能的新型 Illumeo 软件的一个组件，该软件可自动获取相关的放射科先前的检查结果。用户可以在特定的 MPI 视图中点击解剖结构的区域，AI 将查找并打开先前的成像研究以显示相同的解剖结构，切片和方向。对于肿瘤学成像，在图像中点击几次肿瘤，AI 将执行自动量化，然后对先验进行相同的测量，呈现肿瘤评估的并排比较。这可以显着减少与肿瘤跟踪评估和加速工作流程相关的时间。

当放射科医师开始研究时，所有这些信息都以简明的形式呈现，并极大地增强了这位患者的健康状况。爱克发表表示，目标是提高放射科医师对患者的理解，从而改善诊断，治疗以及由此产生的患者结果，而不会增加临床医生的负担。

需要访问大量的患者数据和图像来提供 AI 软件算法教育材料以供学习。通过大量的大数据进行排序是 AI 如何学习临床医生的重要内容，哪些数据元素与各种

疾病状态相关并获得临床理解的重要组成部分。这是一个类似的过程，医学生学习绳索，但使用更多的教育输入比人类可以理解的。机器学习软件的第一步就是要学习医学教科书和护理指南，然后回顾一下临床病例。与人类学生不同的是，AI 用来学习数百万的数字。对于 AI 未能准确判断疾病状态或发现错误或不相关数据的情况，软件程序员在迭代后返回并细化 AI 算法迭代，直到 AI 软件在大多数情况下得到正确的结果。在医学中，变量太多，难以总是对人或机器进行正确的诊断。然而，智慧百分比方面，专家们现在认为人工智能软件阅读医学成像研究往往可以匹配，或在某些情况下，超过人类放射科医生。对于罕见的疾病或表现尤其如此，放射科医生在整个职业生涯中只能看到少数这类病例。人工智能的好处是可以从档案中回顾数百甚至数千次这些罕见的研究，使他们精通阅读并确定正确的诊断。而且，与人类思维不同的是，它始终在电脑的脑海中保持新鲜。人工智能算法通过识别模式来读取类似放射科医生的医学图像。人工智能系统使用大量检查进行训练，以确定来自 CT，磁共振成像（MRI），超声或核成像扫描的正常解剖结构。然后使用异常情况训练 AI 系统的眼睛以识别异常，类似于计算机辅助检测软件（CAD）。然而，与 CAD 只是放射科医生可能想要仔细研究的区域不同，AI 软件具有更多的分析认知能力，基于更多的前几代 CAD 软件的临床数据和阅读体验。出于这个原因，正在帮助开发医学人工智能的专家经常将认知能力称为“有效的 CAD”。

人工智能和放射科的下一步麻省总医院放射科计算和信息科学副主席 Keith Dreyer 博士表示，深度学习计算机已经在驾驶汽车，监测盗窃金融数据，能够翻译语言并识别基于面部识别的人的情绪。波士顿。他是 11 月在北美放射学会（RSNA）开幕会议上的主要发言人之一，他在会上讨论了人工智能进入医学成像领域。他还负责其机构开发自己的 AI 系统，以协助 Mass General 的医生。Dreyer 解释说：“数据科学革命大约在五年前随着 IBM Watson 和 Google Brain 的出现而开始。他说，2012 年推出的深度学习算法确实推动了人工智能的发展，到 2014 年，机器正确读取放射学研究的比例开始下降，准确度达到了 95% 左右。德雷尔说 AI 成像软件并不新鲜，因为大多数人已经在 Facebook 上使用它来使用面部识别算法自动标记朋友的平台身份。他说训练人工智能是一个类似的概念，在这里你可以开始显示一个电脑的照片猫和狗，它可以训练，以确定在使用足够的图像后的差异。Dreyer 说，人工智能需要大量的数据，强大的计算能力，强大的算法，广泛的投资，然后从编程的角度进行大量的翻译和整合，才能被商业化。他从放射学的角度说，有两种类型的 AI。美国食品和药物管理局已经开始批准的第一种类型是定量 AI，只需要 510 (k) 的批准。为临床解释开发的 AI 将需要 FDA 的上市前批准（PMA），涉及临床试验。在机器开始进行初级或同行评审之前，德雷尔说人工智能更有可能被用来回顾旧的检查，以帮助医院找到病人可能没有意识到病情的新病人。他说大约有 900 万美国人有资格接受低剂量的 CT 扫描来筛查他们的肺癌。他表示，人工智能可以接受培训，通过在卫生系统中记录的所有先前的胸

部 CT 检查来帮助识别可能患有肺癌的患者。这种类型的回顾性筛查也可能适用于其他疾病状态，尤其是如果 AI 可以将基因组测试结果拉到狭窄的范围内，使患者易患某些疾病。他说，总的来说，人工智能提供了一个重要的机会来增强和增强放射科的阅读，而不是取代放射科医生。德雷耶说：“我们专注于麦克风说话，而我们忽略了病人记录中的所有其他数据。”“我们需要将影像作为患者的另一个数据来源。”他表示，人工智能可以帮助自动进行鉴定，并迅速从电子病历中提取相关患者数据，以帮助诊断或了解患者的状况。众包治疗选择和监测药物反应

医疗领域的人工智能和机器学习将继续得到改善，影响疾病预防和诊断，通过各种临床试验从数据中提取更多的意义，帮助开发基于个人独特 DNA 的定制药物，并告知治疗选择等等。人工智能辅助再现性：几年前，西门子医疗集团率先将人工智能（AI）算法引入心脏回波系统，以加速自动化。几年前，飞利浦医疗保健公司也在其 Epiq 超声系统中引入了 AI 的元素。它需要一个三维回波数据集采集和自动分析图像，以确定心脏的解剖，标签，然后切片的最佳标准视图呈现。这消除了互操作性差异的问题，因为软件将总是选择基于机器学习的最佳视图，该机器学习使用数千个代表患者解剖变异谱的先前检查。这对于操作人员来说要积累相同的知识需要花费数年的时间。其他供应商也引入了深度学习算法的元素来帮助分析超声心动图或执行自动量化。下一代回声系统将结合更多的人工智能功能，通过自动完成耗时的任务和扩大超声检查员的工作量，从而进一步改善工作流程，从而提高工作效率，始终保持准确。所有主要的成像系统供应商都在开发他们自己的 AI 或与 AI 供应商合作。

1.2 国内外研究现状及难点

1.2.1 医学图像内容理解的研究现状

通用电气，西门子和飞利浦是超声心动图供应商之一，将深度学习算法整合到回声软件中，帮助自动从三维超声数据集提取标准成像视图。这是飞利浦 Epiq 系统的一个例子，该系统使用供应商的解剖智能软件来定义解剖结构，并自动显示解剖标准诊断视图，无需人工干预。这可以大大加快工作流程并减少操作员之间的差异。基于人工智能（AI）的医学图像分析采用卷积神经网络，支持向量机，模糊逻辑系统等机器学习方法从医学图像中提取意义。最先进的计算机视觉软件为诊断人员提供了基于证据的技巧，消除了可能的疑惑并确保了诊断的一致性。标准视图位置是超声心动图中的关键步骤，因为这些帧包含基本的诊断数据。从超声波检查自动捕捉标准飞机可以加快扫描，并使其更加准确。仔细研究这方面的研究将证明这不是一个猜测。标准视图的计算机辅助检测不断支持临床医生。计算机如何看到图像医学图像分析是计算机视觉的实际应用 - 计算机科学的一个分支，涉及数字图像（包括数字视频帧）中的对象和特征识别。计算机视觉算法通过一系列过程

来分析图像，类似于人类视觉系统所执行的过程。在经过初步预处理（包括去噪，滤波和特征增强）之后，软件在图像分割的过程中将图像分解成有意义的区域。然后，算法提取重要的特征，并基于这些特征对图像中的对象进行分类。此外，医学图像分析算法通常执行图像配准 - 映射两个以上相同解剖结构的图像以检测任何差异或变化。基于机器学习，分类是医学图像分析软件最复杂的功能。每个 AI 系统都使用机器学习方法作为其“大脑”。这些算法允许计算机记住大量信息，并在学习完成后使用它来分析类似的信息。这就是为什么这种方法在计算机视觉中得到如此广泛的应用 - 在图像数据集（例如超声图像数据集）上进行训练，然后软件识别真实世界图像中的熟悉特征（例如，在实时超声扫描中）和在此基础上作出相关的结论。这些系统的准确性随着输入数据的数量而增加。从数百个图像开始，它们显示出不错的结果，并且在处理了数以千计的图像和更多图像之后，它们的准确度接近 100%。当然，这也取决于所使用的架构，随着机器学习方法的发展，用于医学图像分析的算法显示出更好的结果。计算机视觉在超声心动图中的应用心脏回声有一些挑战，医学图像分析可以解决。例如，研究人员建议使用计算机视觉自动分割解剖结构，检测和分类先天性心脏缺陷，实时导管定位等。标准视图采集是心脏超声最基本的任务，也可以通过医学图像分析。

1.2.2 人工智能在医学图像领域的原理及研究现状

标准视图获取为了找到标准的心脏视图，软件应该从超声波扫描期间的多个帧中选择合适的二维平面。在这里，出现了不同的挑战，如分析二维帧，三维体积，二维时间序列或四维时空图像相关（STIC）体积。国际合作已经解决了后一个问题，提出了使用尺度不变特征变换（SIFT），最先进的特征检测算法和支持向量机（SVM），监督机器学习方法。此方法已在包含正常和异常情况的数据集上进行了测试。软件定位三个标准平面：四腔切面（A4C），三切面切面（TVV）和横切腹切面（TAS）。该方法在综合数据上显示出了很好的效果，即在随机选择的飞机中检测到了标准的心脏视图，准确率达 87-100%。但是，在实际数据上，表现中等，精确度为 33-53%。不过，在这两个数据集上，该方法比之前的方法显示出更大的结果。未来，研究人员计划通过应用深度卷积神经网络（CNN）来提高准确性，CNN 是用于图像分析的最有前途的机器学习方法。另一个合作提出了一个类似的方法：他们已经应用了基于两个 CNN 的融合深度学习框架来定位三维回波中的八个标准心脏视图，并且达到了 92.1% 的准确度。当仅定位三个主要飞机时，准确度高达 98%。值得注意的是，这两项研究都使用相对较少的数据来训练他们的系统：他们向 SVM 和 CNN 系统输入了对应于数百个超声平面图像的数据。这可能足以测试一个系统的性能，但经过对大数据集的严格培训后，机器学习软件将显示出更好的结果。最新技术和期望今天，CNN 被认为是机器学习中最强大的分类技术。专门设计来分析图像，他们显示壮观的图像分类准确性。在一些任务中，他们已经

超越了人类，正如年度 ImageNet 视觉识别挑战所显示的那样。获奖的 ImageNet 研究团队拥有数百万个标记图像来训练他们的卷积神经网络。因此，随着医学图像数据量的不断增长，我们可以期待医学图像分析软件很快成为超声系统的重要组成部分。

1.3 全文结构及创新点

1.4 本章小结

第二章 医学影像与深度学习相关知识

我们提出了一个完全卷积编码器 - 解码器框架的图像残差变换任务。所提出的框架不是仅使用每像素丢失函数，而是结合依赖于来自预先训练的网络的低级特征的感知损失函数来学习端到端映射。通过引入身份映射指出映射函数以处理无噪声图像。并通过分析神经网络与他们试图学习的基本噪声分布之间的相互作用。我们还展示了如何构建一个统一的变换，然后使用这个统一的变换使单个深度神经网络能够在不同的噪声级别上正常工作。与以前的方法相比，我们的性能更好。实验结果表明了该算法处理图像去噪任务的有效性。

2.1 医学影像

2.2 人工神经网络

2.3 卷积神经网络

深度 CNN 是多层前馈神经网络的一种特例。隐藏层的神经元设计成跟上一层神经元局部连接，并利用参数共享来减少模型复杂度。针对图像这种结构化数据，由不同卷积核来探测不同空间位置上的局部统计特征。通过堆叠多层的卷积结构，实现从低层到高层语义空间的抽象映射。深度 CNN 的典型结构是在 LeNet 模型^[1]的基础上引入修正线性单元 (Rectified Linear Units,ReLU) 的激活函数和 Dropout 等技术^[2] 进行了改进。为 CNN 模型的网络结构示意图。定义图像数据为，且其类别标签，其中 n 和 k 为类别数，作为网络输入，输入层的，即原始图像作为输入，第 l 层输出 n_l 个大小为 (h_l, w_l) 的特征图。第一层为由 n_0 个特征图作为输入的卷积层，特征图大小为 (h_0, w_0) 。第 l 层第 i 个特征图定义为。计算公式为：(1) 其中 $\theta_{l,i}$ 为偏置矩阵， ϕ 为连接第 l 层第 i 个特征图和第 $l-1$ 层第 j 个特征的卷积核。模型的激活函数没有采用 Sigmoid 函数或双曲正切函数，而是选择 ReLU 函数，目的是引入更多非线性来加速训练收敛速度，解决多层网络反向传播中梯度弥散的问题。其函数表达式为：(2) (3)

其中 ϕ 表示对第 l 层的激活函数，该层一般嵌入在卷积层后。为了使得每层输入的分布更平稳，一般引入批量归一化层 (Batch Normalization, BN)，如图 1 中所示。最大池化层进行下采样，有时把“卷积-激活-归一化-池化”统称为卷积层。最后需连接全连接层 (图中 Fc 层表示)，全连接层就不再保存空间信息，是对低层特征的高层抽象，最终输出 K 维的向量，作为该图像的特征向量送入最终的分类器进行分类评估。

2.3.1 全卷积网络

图 1 卷积网络模型结构示意图 Fig.1 The structure of convolutions model 深度 CNN 模型的分类器与传统方法不同的是：把特征提取过程中的卷积核参数和分类器的参数整合到端到端的模型中。对一个有监督的多分类问题，特征提取过程可表示为得分函数， W , b 是各层可学习的参数包括卷积核 K , 偏置 B 和全连接层的权值参数。对第 i 个样本的得分函数分类误差的交叉熵损失函数可定义为：(4) (5) 通过最小化 Softmax 函数的非负对数似然（公式 5），能带来归一化的概率解释。一般采用 L2 损失正则化技术提升分类泛化性能。全部 N 个样本的损失函数 L 为公式 6 所示。其中 λ 表示正则化参数。模型最小化方法采用反向传播算法，通过带动量的批随机梯度下降算法不断调整参数使得模型整体误差函数不断降低。并通过使用权重衰减项和 Dropout 技术控制过拟合。具体实现详情请参考文献 [10]。

2.3.2 全卷积网络

2.4 深度卷积神经网络

2.5 本章小结

第三章 超声心动图切面的识别方法

提出了一种基于深度卷积神经网络自动识别超声心动图标准切面的方法，并可视化分析了深度模型的有效性。该算法针对网络全连接层占有模型大部分参数的缺点，引入空间金字塔均值池化替代全连接层，获得更多的空间结构信息，并大大减少模型参数、降低过拟合风险，通过类别显著性区域将类似注意力机制引入模型可视化过程。通过超声心动图标准切面的识别问题案例，试着对深度卷积神经网络模型的鲁棒性和有效性进行了解释。在超声心动图上的可视化分析实验表明，通过改进方法的深度模型的识别决策依据，同医师辨别分类超声心动图标准切面的依据一致，表明了方法的有效性和实用性。

3.1 引言

在心脏病常规临床检查中，二维实时超声心动图常用于评测心脏的结构和功能。临床超声检查通常主要包括三个步骤：探头扫描不同位置，选取标准切面和对标准切面的测量和诊断^[3]。其中，医师总结出来能更好辅助分析心脏功能结构的特定位置和角度的超声心动图称为标准切面，其正确快速选取不仅对临床诊断具有至关重要的意义，也为病例研究提供比较依据。标准切面的自动识别是超声心动图智能分析和测量的基础。与自然图像相比，医学超声成像质量差，存在斑点噪声和伪影；并且各标准切面类内、类间差异大，使得标准切面的识别成为一个非常具有挑战性的问题。

目前的研究主要集中在利用机器学习和图像处理等方法，进行超声心动图的自动识别、检索及切面内组织结构的定位和分割等。针对超声心动图的自动识别，2004 年 Shahram 等^[4]首次提出采用马尔科夫随机场，设计通用腔室模板检测心脏腔室来辅助三类标准切面识别，但需额外信号来指定处于舒张末期（End-Diastolic, ED）的切面。同样利用处于 ED 的标准切面，Kevin 等^[5]基于多类别提升算法框架，提取哈尔矩形特征训练弱分类器，同样需要检测心脏腔室的空间位置，辅助四类标准切面识别。基于降低特征维度的两层级联方法，把标准切面分类成心尖和胸骨旁两大类，然后进一步区分四类标准切面视频^[6,7]。在文献 [4] 工作基础上整合局部和全局模板特征，利用多类逻辑提升分类算法，并指出能扩展到任意标准切面^[8]。在对心脏的循环跳动的时空信息进行统计分析的基础上，利用主动外观模型对形状和纹理进行建模，统计追踪一个心动周期并投影到运动空间进行分类^[9]，该方法处理的视频序列。把标准切面视为不同场景，提取低层全局特征来表征不同切面，利用改进核支持向量机进行分类^[10]。这些方法可以归纳为两个阶段：首先根据先验人为设计特征来表征图像；然后利用机器学习中不同分类方法对特征向

量进行建模分析得到分类器。然而受限于‘语义鸿沟’问题，根据特定先验人为设计特征，如大多数方法都针对心动周期的某个特定时刻的切面（如 ED），会导致模型泛化性能差。

近来，深度卷积神经网络 (Convolutional Neural Network,CNN) 在大规模自然图像数据集 (如 ImageNet^[11]) 上，识别性能远超传统方法^[2]。主要得益于深度学习利用大量标注数据从图像原始像素出发，逐层分级学习中高层的抽象语义特征^[12]。当前实践中由于深度学习需要大量的标注数据，所以仅在少数医学任务中取得有限的成功应用，且对深度模型的鲁棒性和有效性也缺乏详尽分析。Chen 等 [3] 利用 CNN 结合领域知识，在胎儿超声心动图标准切面的自动识别问题中取得良好的识别效果，但胎儿跟成人超声心动图差异大，具有很大特殊性。Bar 等^[13] 利用自然图像训练的模型对胸腔 X-射线图像进行特征提取并结合全局特征^[14] 得到最优检测结果，并没有对特定医学数据进行迁移训练，仅是作为特征提取器。Margeta 等^[15] 针对心脏核磁共振图像利用微调迁移从自然图像学习的模型，但没对模型有效性进行分析。

目前深度 CNN 模型的理论分析工作还不是很完善，能自动学习语义特征的工作机理还是个“黑箱”。对于不同的模型的比较除了准确率之外并没有很好的评价方法，优异的泛化能力从何而来仍是个开放问题。一些工作^[16-19] 通过可视化各层激活值和卷积核来更好理解深度 CNN。对在给定数据集上训练得到的深度 CNN 网络模型，Simonyan 等^[16] 用反卷积来可视化每个神经元的最大激活值。Mahendran 等^[17] 通过对学习到的每层的特征编码进行反编码，建立每层特征编码和原图像的映射关系。Zeiler 等^[18] 试图通过梯度上升方法迭代寻找图像使得最大化激活某个或某些特定的神经元。神经元对图像每个像素的梯度描述了当前像素的怎样改变能影响分类结果。前三个方法均是对已训练的模型进行分析，而类激活映射图 (Class Activation Maps,CAM) 方法^[19] 用全局平均池化层代替全连接层改进训练过程，分类性能虽略有降低，但能指示出特定类别的显著性判别区域，能很好的解释模型的有效性。本文提出一种基于深度 CNN 识别超声心动图的方法 (Deep Echocardiogram,Deep-Echo) :1) 引入空间金字塔平均池化层代替全连接层，一方面大大减少模型参数，降低过拟合风险；另一方面网络结构变为全卷积网络，使得不用限制输入图像尺寸大小，这对医学超声图像更为重要。2) 为验证该算法的鲁棒性和有效性，针对数据集进行详尽实验，研究分析了深度学习方法的高识别率和优异泛化能力的原因。

3.2 Deep-Echo 模型

将分别从如何构建全卷积网络、全局空间金字塔平均池化层、将类别显著性图纳入可视化过程、如何扩增数据等方面介绍提出的 Deep-Echo 模型。

3.2.1 全卷积的网络

与 GoogLeNet 模型^[20]、ResNet 模型^[21]类似，使用多层卷积层（每层包括 ReLU 层、BN 层和 Pooling 层），用全局平均池化操作替代全连接层。Deep-Echo 模型结构中对最后卷积层输出的特征图，用金字塔平均池化层^[22]代替最大化池化层和全连接层。最后一层输出单元数目为类别的数目，由于实验采用的标准切面有七个类别，因此最后一层输出 7，依次对应相应的类别，采用交叉熵损失函数加 L2 正则化。卷积核数目从 64 开始，每经过一次最大池化层，卷积核数目翻倍，直到 512 为止。学习率初始化为 0.01。具体实验步骤和参数设置见后文实验部分。整个网络结构如图3.1所示。

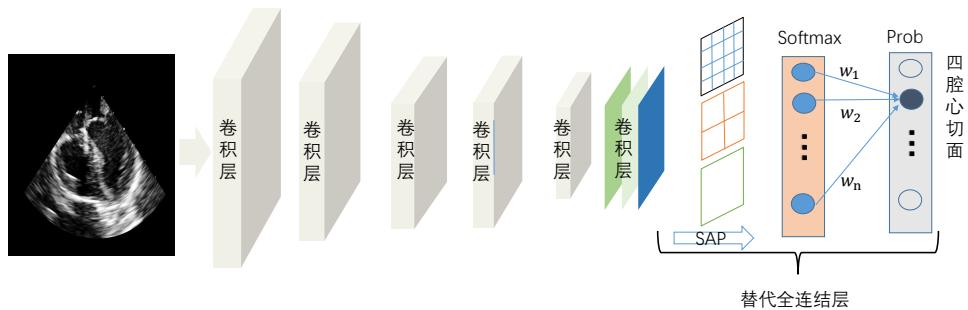


图 3.1: Deep-Echo 模型结构示意图

3.2.2 空间金字塔均值池化层

针对深度 CNN 模型中全连接层的两个缺点：全连接层丢失了空间信息，限制了 CNN 只能接受固定尺度的输入，一般只能通过图像尺度归一化的方法来处理不同尺度的输入图像，且使得模型可视化变得不可解释；全连接层参数拥有大约 90% 的模型参数，如 AlexNet 模型^[2] 和 VGG16 模型^[23] 中全连接层参数占全部参数分别为 38M/61M 和 103M/138M，从而导致模型更容易过拟合^[20]。为解决这两个问题，He 等提出空间金字塔池化 (Spatial Pyramid Pooling, SPP) 方法^[22]。SPP 通过使用多个不同大小的池化操作保证固定的特征向量输出，从而允许 CNN 接受任何尺度的输入，增加了模型的尺度不变性，抑制过拟合。与传统的全连接层不同，对每个特征图一整张图片进行多尺度的空间金字塔均值池化，这样每张特征图都可以得到多个尺度的输出。本文方法跟空间金字塔池化网络类似都是三个尺度的空间金字塔池化 (1×1 , 2×2 , 4×4)，其差异在于后不再接多个全连接层，同时用平均池化代替最大化池化，目的在于方便可视化模型的空间位置信息。

3.2.3 微调迁移学习

利用深度学习进行超声心动图的标准切面识别，仍存在针对小数据量直接训练是否会出现过拟合问题；能否跨领域进行迁移学习，即在自然图像数据集上训练得到的模型能否微调应用到跨领域的超声心动图上。文献 [19] 中指出，用全局平均池化代替全连接层直接随机初始化，从头开始训练模型收敛困难且分类性能下降，故对现有模型进行改造，即针对在自然图像集上预先训练得到的模型如，Alexnet 模型等，变换最后的输出层为所述金字塔平均池化结构，调小学习率后在超声心动图标准切面数据上进行微调迁移学习。训练时，由于超声心动图的特殊性，人工标注费时费力，对数据集进行扩增能降低人工标注的需求。但扩增数据需注意不能打乱标准切面图像内在的局部结构，因此对切面数据只进行水平镜像翻转和旋转。通过引入 BN 归一化层能减轻对 Dropout 的依赖，提高泛化能力，并且本文直接去掉全连接层，故并未采用 Dropout 技术。迁移学习时，由于深度模型中低层的卷积核是跟人类视觉的初级细胞很类似，因此是可以直接迁移复用，高层要针对目标学习判别性信息需进行重新学习^[19]。针对超声心动图的实验支持这样的结论，不同模型的分类准确率都很高，具体实验见后文实验部分。但对于计算机医学辅助诊断而言，模型怎样决策判断比分类准确率更重要。即需解释模型为什么有效和优异的泛化能力从何而来。

3.2.4 类别显著激活映射图

前文所提模型能高效提取超声心动图标准切面的特征，对超声心动图的单扇形和双扇形标准切面都能很好的识别，甚至对互联网上随意选取的标准切面也能识别。但对模型的有效性和解释性缺乏有力分析，使得对模型决策判断的可信性产生怀疑。针对超声心动图，采用 [19] 提出可视化分析的方法，将其和空间金字塔平均池化结合。对给定图像， $f_j(x, y)$ 表示卷积层 (x, y) 位置上第 j 个神经元的激活值，对第 j 神经元的平均池化操作结果对给定类别 k 的得分函数 S ：

$$S_k = \sum_j w_j^k \sum_{x,y} f_j(x, y) \quad (3.1)$$

其中 w_j^k 是第 j 个神经元和第 k 类的连接权重，后接多类多元逻辑损失层，然后由公式3.2可得定义类别激活映射图：

$$M_k = \sum_j w_j^k f_j(x, y) \quad (3.2)$$

其中， M_k 表明在空间 (x, y) 的激活值对该类别分类结果影响的重要性。对类别激活映射图直接双线性插值得到与原图大小相等的显著性图。本文将其和多尺度空间金字塔平均池化结合，得到对多个空间尺度的类别显著激活映射图。值得注

意的是，对不同的尺度可设置不同的权重，本文采用同等权重进行融合。该图是对图像空间显著性区域的置信度判别，能辅助可视化分析深度模型的决策过程，在一定程度上解释模型可效性。

3.3 实验结果和分析

3.3.1 实验数据选取和实验方法

本文实验数据来自四川大学华西医院，为临床检查中的经食道超声心动图。所选切面视频包含单扇形和多普勒成像的双扇形两种，其中对双扇形的切面视频，仅取不包含彩色多普勒成像的切面（如图3.2所示）。经专业医师标注的标准切面视频中，至少包含2-3个心动周期，并依据医师建议从视频中截取包含一个心动周期的10帧图像，并经医师检验筛选后得到最终数据集。

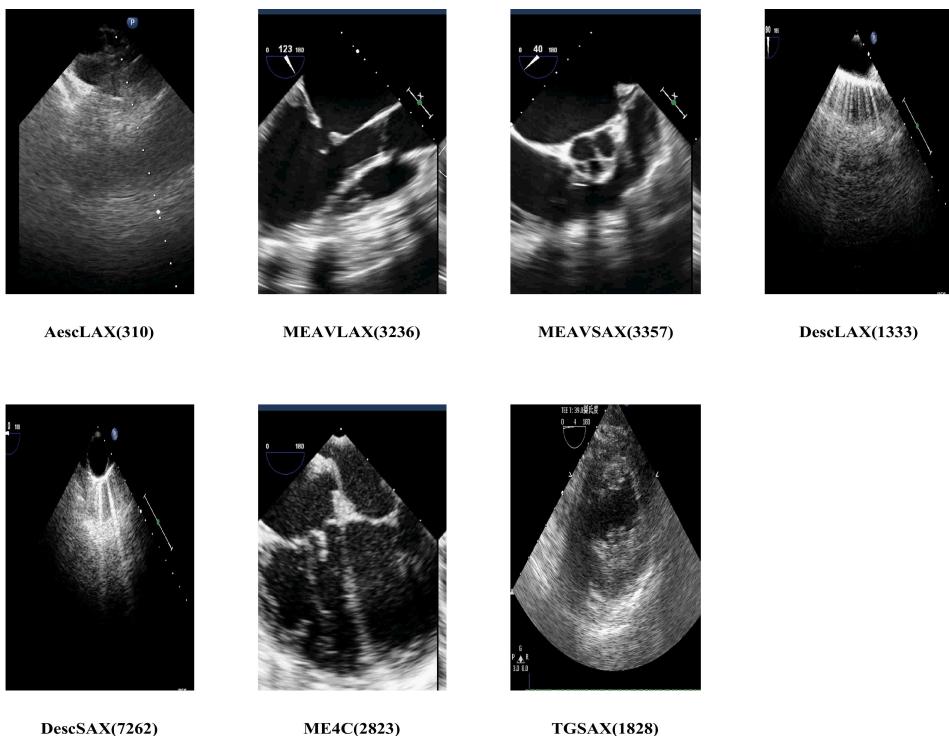


图 3.2: 七类标准切面超声心动图及数量分布

试验中所用标准切面类别和数量分如布图3.2所示。依据探头在食管中段(ME)和经胃底(TG)的位置和角度不同，在图3.2中7类标准切面分别为：a为升主动脉长轴(AescLAX)，b为主动脉瓣长轴(MEAVLAX)，c为主动脉瓣短轴(MEAVSAX)，d为降主动脉长轴(descLAX)，e为降主动脉短轴(descSAX)，f为食管中段四腔心(ME4C)，g为经胃底心室短轴(TGLAX)。其中，d，e，g为单扇形切面，其余为双扇形中截取的切面。训练集(17932张)和测试集(2217张)由不同时期采集

不同病人对象数据的随机划分。值得注意的是，所有数据都经过裁剪操作以隐去患者信息。

3.3.2 识别实验结果和分析

本文在构建的超声心动图的数据集上测试分类性能。采用 Caffe 框架^[24] 实现深度卷积网络结构，预训练模型来自 Caffe model zoo。使用具有 Intel®Core™ i5 3.2GHz 处理器和 12GB 内存的 Tian X GPU 测量所需的时间，单个切面所需的分类识别时间平均需要 10 毫秒，基本可满足实时识别。为验证从自然图像训练的模型能迁移到经食道超声心动图上，输入图像归一化为 256x256，网络初始学习率设为 0.001，迭代一定轮数动态调整学习率大小，其他参数的设置跟原文献中训练网络结构时一致。三种不同网络结构的深度模型微调前后在同一测试集上的准确率随着迭代次数的增加最后趋于一致，如表3.1所示，Scratch 表示不经过微调，Finetune 表示经过微调。Deep-echo 模型结构跟 AlexNet 模型类似，是在其结构基础上去掉全连接层，用空间金字塔池化层代替，比 VGG16 和 GoogLeNet 模型的层数更少，模型结构更简单，而分类准确率却接近，表明提出方法的有效性。针对 VGG16 模型和 Google Net 模型也可同样设置，本文主要关注点不是得到分类精度最优的分类模型，故并未全部加以实验验证。为验证训练集数据量对深度卷积网

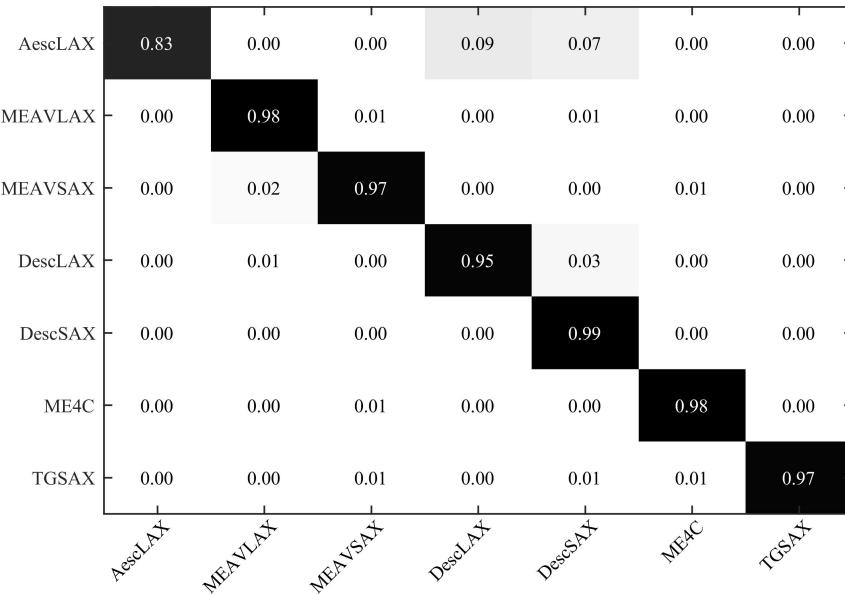


图 3.3: 不同数据量的平均分类精度

络的影响。网络结构采用 AlexNet 模型结合空间金字塔池化层，在不同数据量上微调，实验结果如图3.3所示，数字代表每类至多的数目，随着数据量的增加，模型准确率随之提升，可知针对超声心动图标准切面识别问题，并不用构建很大的数据集进行识别，如图3.2中每类至多 500 达到的平均准确率接近使用全部训练集的结果。可推断采用微调技术，能显著减少深度模型对大数据量的依赖。

平均分类精度比较		
	Scratch	Finetune
AlexNet	93.35%	93.68%
VGG16	96.66%	96.81%
GoogleNet	97.36%	97.42%
Deep-Echo	97.49%	99.12%

表 3.1: 不同模型分类精度比较

为了验证最优模型在不同类别的分类性能, 7 分类的混淆矩阵如图3.3所示, 每行代表实际的类别标签, 每列代表预测的标签。最终的平均分类精度为 97.49%。分类置信度较低的是升主动脉长轴 (AescLAX), 其他各类的准确率都较高。

3.3.3 模型可解释性实验结果分析

深度卷积网络能在标准切面识别问题上得到较高的分类精度, 但仅从分类准确率上评价模型存在局限性。为分析模型的有效性, 采用文中所述可视化方法, 对迁移后的 Deep-echo 模型进行实验。

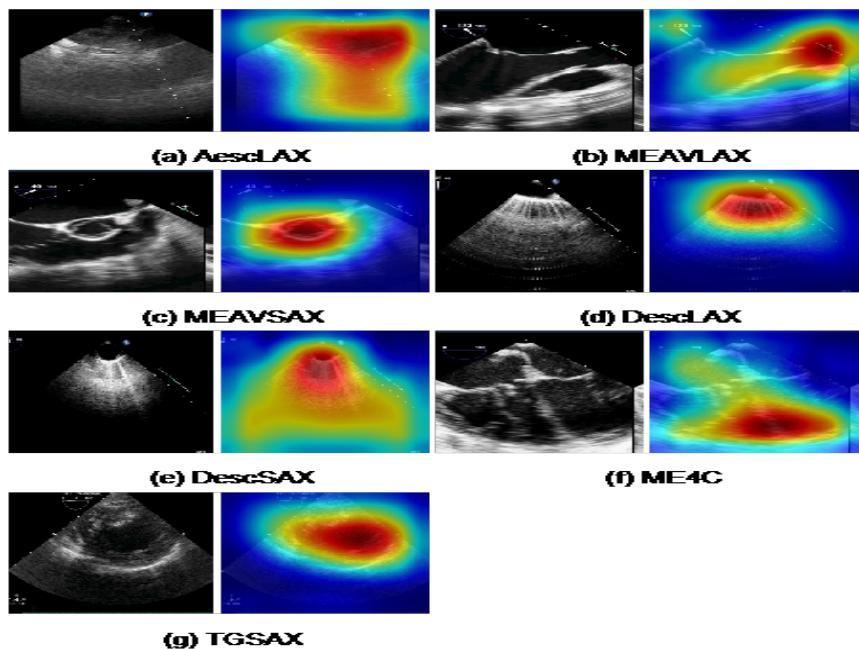


图 3.4: 各类切面的原图和显著性热力图

实验结果如图3.4各类切面的原图和显著性热力图所示, 图中为各类切面和对应的类别显著性图中的颜色从蓝到红, 表示原图像素中对分类

结果影响的重要性是从轻到重。图中结果能很好的解释模型的有效性，并且跟专业医师的判断一致，如图3.4c 中显著性热力图红色区域图定位到图中的圆圈；图3.4d 中定位到的干涉条纹；图3.4f 定位到左心室和右心室的边界等；都跟医师的决策判断依据是一致的。

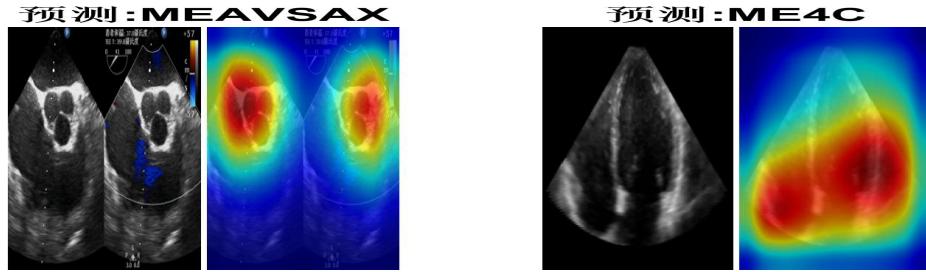


图 3.5: 深度模型泛化性能可视化分析

深度模型泛化性能可视化结果如图3.5所示，原图像分别是带彩色多普勒的双扇形切面（图3.5a）和经胸的四腔心切面（图3.5b），这两个图是跟数据集中的经食道超声心动图差异较大，说明深度卷积网络模型确实能对标准切面进行语义分类，表明模型确实能提取到高层语义的特征，深度卷积网络泛化能力优异。图 7 中可视化结果也能很好的解释模型的有效性，如图3.5中显著性热力图红色区域图定位到图中的圆圈，也是医师认定该切面的关键性结构，图3.5b 定位到左心室和右心室的边界等，都跟医师的决策判断依据是一致的。并且该方法也能作为判断学习模型是否有效的根据，不经过微调的模型虽然能得到较高的分类准确度，并不能得到类似的显著性热力图。

3.4 本章小结

本文提出了一种基于深度卷积神经网络的超声心动图标准切面自动识别方法，利用所述全局空间金字塔均值池化方法进行微调迁移学习，实验结果表明该方法识别准确率高，并实验分析了数据规模对模型分类精度的影响，结果表明基于深度卷积网络的识别方法应成为超声心动图自动识别的基准方法，接下来会探索更精细类别分类问题，如舒张末期和收缩末期标准切面的识别等。可视化深度模型的实验，对模型的可解释性和有效性进行了分析，推断深度模型的优异的分类性能和泛化能力的原因是可以对类别显著性区域进行判别，采用的可视化方法是对网络模型整体的理解，具体各层特征怎么耦合成语义信息仍需进一步探索。

第四章 空间金字塔分解的深度可视化方法

提出了一种基于深度卷积神经网络自动识别超声心动图标准切面的方法，并可视化分析了深度模型的有效性。该算法针对网络全连接层占有模型大部分参数的缺点，引入空间金字塔均值池化替代全连接层，获得更多的空间结构信息，并大大减少模型参数、降低过拟合风险，通过类别显著性区域将类似注意力机制引入模型可视化过程。通过超声心动图标准切面的识别问题案例，试着对深度卷积神经网络模型的鲁棒性和有效性进行了解释。在超声心动图上的可视化分析实验表明，通过改进方法的深度模型的识别决策依据，同医师辨别分类超声心动图标准切面的依据一致，表明了方法的有效性和实用性。

4.1 引言

以深度卷积神经网络（Convolutional Neural Network，CNN）为代表的深度学习对计算机视觉和机器学习领域产生了深远影响。但是完全理解深度学习模型的内在工作原理，设计高性能的深度网络结构还是很困难的，一直以来人们普遍将其内部工作原理看成一个“黑箱”，这是由于深度 CNN 存在海量参数，多次迭代更新生成输入输出之间相当不连续和非线性的映射函数；以及对参数的初始状态敏感，存在很多局部最优点。探究 CNN 的运行机制，核心在于它究竟自动提取什么样的特征，经过卷积层、池化层，特征都是分布式表达的，每个特征反映在原图上都会有重叠，故希望建立特征图与原图像之间的联系，即深度可视化。该技术试图寻找深度模型所提取各层特征较好的定性解释，并在设计开发新网络结构方面扮演重要角色。

目前针对 CNN 可视化的研究，主要集中在如何理解 CNN 从海量数据中自动学习到的，能反映图像本质的分层特征表达，即获得网络中隐藏层神经元与人类可解释性概念之间的联系。最直接的方法是展示学习得到的卷积核和相应的特征图，但除了首层卷积核和特征图有直观的解释外，其余各层并没有可解释性。从信号处理的角度看，基于 CNN 高层特征的分类器在输入域，需要较大感知野，才能对以由低频为主的输入图像进行多层非线性响应，并对小的输入改变产生平滑不变输出。同时，由于经过非线性激活函数变换和池化，引入空间不变性获得更好识别性能的同时，也对可视化带来新的挑战。

深度可视化技术可以简单分为三类：基于梯度更新的方法^[16,25~30]；基于特征重建的方法^[17,18,31,32]；基于相关性的方法^[33,34]。基于网络梯度更新的思想是由 Erhan 等^[25]引入，固定模型参数通过梯度更新改变输入值，最大化激活单一神经元或标签类别概率。激活最大化生成的非自然图像还可以是网络模型的对抗样本^[35]。

Simonyan 等^[16,26,27] 通过梯度上升方法迭代寻找使得最大化激活 CNN 某个或某些特定的神经元的最优图像，其假设神经元对像素的梯度描述了当前像素的改变能影响分类结果的强度。文献 [16] 引入 L2 正则化先验（或称权重衰减），改进可视化效果。Yosinski 等^[28] 进一步提出高斯模糊正则化、梯度剪切等技术，其中梯度剪切指的是每次只更新对分类最有利的一部分梯度，改善生成图像质量。文献 [26, 30] 考虑神经元的多面性和利用生成网络作为自然图像的先验来合成更自然的图像。Zeiler 等^[18] 提出利用反卷积网络，利用反向传播重构各层特征到像素空间的映射，并用于指导设计调优网络结构，提高分类识别精度。在反卷积过程中利用翻转原卷积核近似作为反卷积核，针对特定特征图在训练集上重新训练。Dosovitskiy 等^[31] 提出通过学习‘上’卷积网络来重建 CNN 各层的特征，指出结合强先验，即使用于分类的高层激活特征也包含颜色和轮廓信息。Mahendran 等^[17,32] 通过对学习到的每层特征表达进行反编码重建，提出利用全变分正则化和自然图像先验，并将 L2 范数正则化推广到 p 范数正则化，得到较优的可视化效果。

本文主要关注前两种方法中的正则化技术，基于相关性分解方法请参考文献 [34]。受文献^[36,37] 启发，把用于图像生成的拉普拉斯金字塔，进一步扩展成空间金字塔分解方法，并引入显著性激活图技术进一步改进深度 CNN 的可视化效果。

4.2 可视化方法的数学模型

激活最大化和特征表达反编码重建均是针对已经训练好的模型，对给定输入 $x_i \in R^{C \times H \times W}$ ，其中 C 为颜色通道数，H，W 为图像高和宽。CNN 模型可抽象为函数 $\phi: R^{C \times H \times W} \rightarrow R^d$ ，其第 i 个神经元的激活值为 $\phi_i(x)$ ，对给定图像 x_0 的特征编码 $\phi_0 = \phi(x_0)$ ，定义参数 θ 的正则化项 $R_\theta(x)$ ，寻找使得能量泛函最小化的初始输入 x^* ，其数学模型为

$$x^* = \underset{x}{\operatorname{argmin}}(l(\phi(x), \phi_0)) + \lambda R_\theta(x) \quad (4.1)$$

其中， l 损失比较的是 $\phi(x)$ 和目标 ϕ_0 的差异，选择不同的损失函数定义不同的可视化方法。但该优化通常是一个非凸优化问题，通常采用梯度下降法去寻找局部最优值为

$$x \leftarrow x + \alpha \frac{\partial \phi_i(x)}{\partial x} \quad (4.2)$$

激活最大化方法是文献 [25] 中提出针对深度架构中任意层中的任意神经元所提取的特征，寻找使一个给定的隐含层单元的响应值 $\phi_0 \in R^d$ 最大的输入模式，可由内积形式定义 l 损失为

$$l(\phi(x), \phi_0) = - < \phi(x), \phi_0 > \quad (4.3)$$

式中 ϕ_0 需人工指定，最大化激活的目标可以是全连接层的特征向量，也可以是卷积层某一通道的某一神经元的激活值。特征表达的反编码重建，通过最小化给定特征向量与重建目标图像特征向量间的损失，一般采用欧式距离来衡量损失误差，定义如下

$$l(\phi(x), \phi_0) = \frac{\|\phi(x) - \phi_0\|^2}{\|\phi_0\|^2} \quad (4.4)$$

但也可利用其它距离度量函数来评价损失。

4.3 梯度更新的可视化方法

用于分类的深度 CNN 提取高层语义信息的同时，丢失了大量低层结构信息。由于首层卷积核大都类似 Gabor 滤波器，导致梯度更新可视化生成图像中包含许多高频信息，虽然能产生大的响应激活值，但对可视化来说导致生成的图像是不自然的。还由于网络模型的线性操作（如卷积）导致对抗样本^[35] 的存在，为得到更类似真实自然图像的可视化结果，需在优化目标函数中引入正则化作为先验。

4.3.1 p 范数正则化方法

对图像来说，像素大小需在一定范围内，直接最大化激活类别概率，生成图像类似随机噪声图像。文献 [16] 通常引入 L2 范数正则化，惩罚过大和过小的极端值，其公式为 $R_\theta(x) = \|x\|_2^2$ 。在文献^[17] 中将其扩展到彩色图像 RGB 通道空间中的 p 范数正则化为

$$R_\theta(x) = \frac{1}{HWC^p} \sum_{h=1}^H \sum_{w=1}^W \left(\sum_{c=1}^C x(h, w, c)^2 \right)^{\frac{p}{2}} \quad (4.5)$$

式中 h, w 表示图像的行和列大小， c 表示颜色通道数，对比发现，文献 [16] 提出的 L2 正则化是忽视各颜色通道的差异的，正则化的力度可通过缩放常量 p 进行控制，即使得图像像素值大小保持在合适的范围内。

4.3.2 高斯模糊和 TV 变分

基于梯度更新可视化方法，引入高斯滤波器主动惩罚高频信息^[28]，高斯模糊核半径大小由高斯函数的标准差控制，可随迭代次数动态调整模糊核大小。

全变分^[17](Total Variance, TV) 跟高斯模糊类似，鼓励可视化生成分片的常量块区域，对离散图像全变分操作可由有限差分来近似求解为

$$R_{TV}(x) = \frac{1}{HWC^\beta} \sum_{hwc} ((x(h, w + 1, c) - x(h, w, c))^2 + (x(h + 1, w, c) - x(h, w, c))^2)^{\frac{\beta}{2}} \quad (4.6)$$

式中 $\beta = 1$, 但其在可视化过程中, 在图像的平坦区域并不存在边缘, 全变分操作仍沿着边缘方向扩散就会导致出现虚假的边缘, 会引入所谓的“阶梯效应”现象。 $\beta < 1$ 时结合超拉普拉斯先验^[38] 能更好匹配自然图像的梯度统计分布, 但对可视化来说反而使得可视化更困难。文献 [17] 实际实验表明, 跟高斯模糊核一样, 需随迭代次数动态调整 β 大小。

4.3.3 基于数据统计先验

由于常规可视化方法并没有对颜色分布进行建模, 文献^[26] 提出通过引入外部自然图像数据, 计算图像色块先验为

$$R_\theta(x) = \sum_p \|x_p - D_p\|_2^2 \quad (4.7)$$

式中 p 为块索引, x_p 表示稠密采样的归一化图像块, D_p 表示自然图像块数据库中距离 x_p 最近图像块。该方法跟文献^[36] 中利用参考图像“指导”人脸图像嵌入重建类似。并且基于数据的统计先验可进一步扩展, 引入生成对抗网络, 利用生成网络主动生成自然图像先验^[29]。

4.4 空间金字塔分解

前文介绍的正则化先验主动限制图像空间中高频率和高振幅信息, 生成的可视化图像存在如下问题: 1) 彩色图像的颜色分布仍是不自然的。2) 生成的图像中包含可识别类别对象的多个重复成分, 并且这些部件不能组合成完整的有意义整体。3) 缺乏令人可信的低频细节, 存在棋盘效应, 只是形似。针对这些问题提出利用空间金字塔分解, 主动提升低频信息和调控高频信息以改善生成图像的可视化效果。

4.4.1 高斯和拉普拉斯金字塔分解

拉普拉斯金字塔 (Laplacian Pyramid, LP)^[39] 是由一系列包含带通滤波器在尺度可变的图像上加低频残差组成的。首先通过高斯平滑和亚采样获得多尺度图像, 即第 K 层图像通过高斯模糊、下采样就可获得 $K+1$ 层, 反复迭代多次构建高斯金字塔 (Gaussian Pyramid, GP)。用高斯金字塔的 K 层图像减去其第 $K+1$ 层图像上采样并高斯卷积之后的预测图像, 得到一系列的差值图像即为拉普拉斯金字塔分解图像。拉普拉斯金字塔分解过程 (见图 1 所示) 包括 4 个步骤: 1) 高斯平滑; 2) 降采样 (减小尺寸); 3) 上采样并高斯卷积 (图中 expand 操作); 4) 带通滤波 (图像相减)。拉普拉斯金字塔突出图像中的低频分量, 拉普拉斯金字塔分解的目的是将源图像分解到不同的空间频带上。

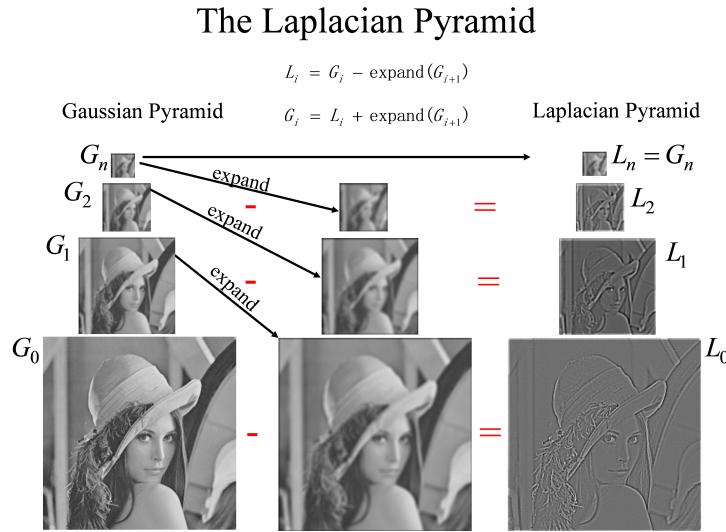


图 4.1: 高斯和拉普拉斯金字塔

由于自然图像统计特性中的尺度不变性，也称为 $1/f$ 法则^[40]，即自然图像集 $I(f_x, f_y)$ 的平均傅里叶功谱服从 $I(f_x, f_y)^2$ 。在激活最大化可视化深度 CNN 模型过程中利用提出的高斯和拉普拉斯空间金字塔分解，调整生成梯度图像包含的频谱分量大小。其中空间金字塔分解正则化项为

$$r_\theta(x) = \sum_{k=1}^K [LP_k + GP_k] \quad (4.8)$$

式中 k 代表构建 k 层金字塔分解，本文实验 k 选取为 4。 LP_k 为第 k 层的拉普拉斯金字塔分量， GP_k 为第 k 层的高斯金字塔分量。

4.4.2 梯度归一化

基于梯度更新的可视化方法，由于原输入空间中高低频分量混杂在一起，对原输入图像相应的更新梯度进行归一化操作能得到较好可视化效果，即对输入图像每次迭代更新的梯度，则提出梯度归一化操作：

$$g \rightarrow \frac{g}{g.std() + \delta} \quad (4.9)$$

式中 δ 为非负小常量， std 表示梯度矩阵的方差。该梯度中心归一化技术，可以减少产生重复的对象碎片的倾向，而倾向于产生一个相对完整对象。梯度归一化的引入同批归一化 (Batch Normalization) 思想类似，校正 CNN 网络非线性变换引起的“偏移”，该方法也侧面验证最新提出的分层归一化^[41] 的有效性。

4.4.3 类别激活图限制可视化区域

根据文献 [19] 提出的类别激活图技术，假设 $f_j(x, y)$ 表示最后的卷积层空间 (x, y) 位置上第 j 个神经元的激活值，则对 j 神经元的全局平均池化操作结果对给定类别 k 的得分函数 S_k :

$$S_k = \sum_j w_j^k \sum_{x,y} f_j(x, y) \quad (4.10)$$

式中 w_j^k 是第 j 个神经元和第 k 类的连接权重。根据文献 [19]，由式 4.10 可得定义类别激活图 M_k 为

$$M_k = \sum_j w_j^k f_j(x, y) \quad (4.11)$$

式中 M_k 表明在空间 (x, y) 位置的激活值对分类结果影响的重要性。对类别激活映射图直接双线性插值得到与原输入图像大小相等的显著性图。本文利用显著性激活图作为梯度更新的权重因子，即输入变为原始输入图像与类别激活图的加权乘积。动机是要求网络梯度更新保持在类别显著性区域内，压制无关背景信息的生成。具体详情请参见第四章实验部分。

4.4.4 优化方法

深度 CNN 模型优化策略的核心是随机梯度下降法，常用方法是带动量的随机梯度下降法为：

$$V_t = \mu V_{t-1} - \alpha * \nabla f(x_i) \quad (4.12)$$

$$x_{t+1} = x_t + V_t \quad (4.13)$$

式中 μ 为动量因子表示保持原更新方向的大小，一般选取 0.9， x_t 为在 t 时刻待更新的梯度， α 为学习率；文献^[17,32] 采用自适应梯度 (Adaptive Gradient, AdaGrad)^[42] 的变种算法，根据历史梯度信息自适应调整学习率。同时文献 [43] 采用的二阶优化算法针对纹理和艺术风格重建问题，得到比用基于一阶随机梯度下降算法更优的可视化效果。但本文通过实验对比发现对各种优化方法对生成图像质量影响不大，从简选择带动量的随机梯度优化方法。

4.5 实验结果分析和讨论

基于梯度更新的可视化方法主要用于激活最大化和特征重建，但文献 [44] 指出用随机未训练的 CNN 模型也能较好重建原图像，表明特征编码重建不能很好解释训练得到 CNN 模型的内在工作机理。故本文实验主要关注在对 ImageNet 公开数据集上预先训练得到的分类模型进行激活最大化可视化实验。

4.5.0.1 不同深度模型的类别可视化

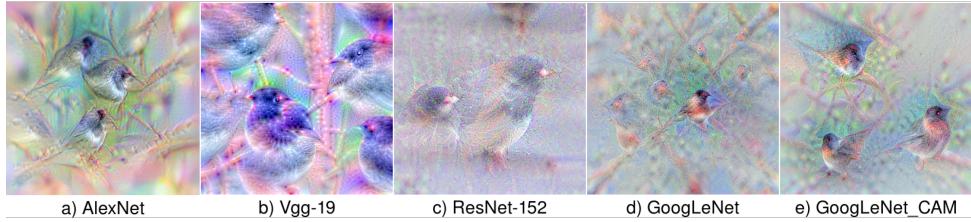


图 4.2: 不同模型类别可视化实验结果

实验选取的深度模型来自于开源社区的 Caffe model zoo，不同的 CNN 模型如：AlexNet 模型^[2]，Vgg-19 模型^[45]，Google-CAM 模型^[19]，GoogleNet 模型^[20]，ResNet 模型^[21]，其分类识别性能依次从低到高，模型的复杂程度依次递增。本文实验默认采用提出的梯度归一化，并引入多分辨率、随机扰动和剪切等小技巧作为通用设置，提高可视化效果。

为比较不同深度 CNN 模型学习相同类别时特征图的差异，根据式4.1，给定高斯噪声生成随机图像作为输入，指定可视化物体类别向量（见图4.2所示，类别为所有类别中的第 13 类布谷鸟），施加前文提出不同正则化项的组合： p 范数、高斯模糊和金字塔分解正则化。

图4.2结果表示 5 种 CNN 模型在相同正则化方法和相同梯度更新策略下的可视化效果，对比图4.2中 a, b, c 发现随着网络模型深度的增加，可视化难度增大分类性能同可视化效果一致；Vgg-19 模型由于跟 ResNet 模型卷积核大小类似，且比 AlexNet 首层卷积核小（7 和 3），即可视化效果倾向生成比 AlexNet 更大尺寸的物体。而由图4.2中 a, d, e 对比可知，由于 GoogleNet 模型中卷积层的卷积核大小不一，使得可视化结果中引入更多细节。综合可知，基于 GoogleNet 模型的可视化效果最好，后面实验均是在其模型的基础上进行实验比较。

4.5.0.2 不同正则化方法的类别可视化

为验证不同正则化方法对理解深度模型的特征表达的影响，采取前文所述的不同正则化方法，可视化效果结果见图4.3所示，从上到下依次可视化类别为金甲虫，海星，蝎子，酒壶，卷笔刀。

图4.3中 (a) 列仅施加默认设置和不加梯度归一化的结果，由于输入的随机性，并不能保证每次都生成有意义的可视化结果，但引入本文提出的梯度归一化后，能大概率生成可视化结果见图4.3(b) 列所示，图4.3(c) 列表示只采用 p 范数正则化，跟文献^[?]^[2] 一致取 2，使得图像更平滑，但仍与真实图像相差较大。通过前文理论

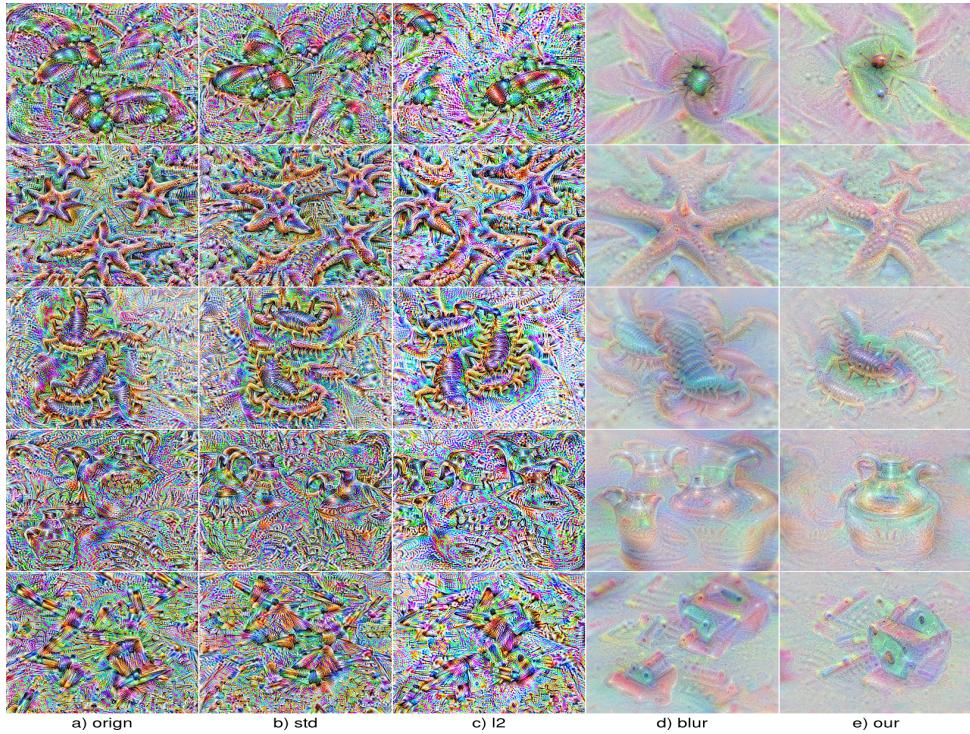


图 4.3: 不同正则化方法的可视化效果

分析和实验验证，全变分跟高斯模糊作用类似，本文采用根据迭代轮数动态调整高斯模糊核大小，具体是在刚开始采用较大值希望生成物体大概轮廓，随迭代逐渐调小模糊核使得更多细节生成，具体见图4.3(d)。但是这个参数无法自适应设置为最优，对图像高低频分量无法调整控制，而本文提出的利用金字塔分解正则化方法能从粗到细调整，产生较优结果见图4.3(e)列所示。

4.5.1 金字塔分解可视化实验结果

为验证提出金字塔分解正则化方法，对中间层卷积核的可视化，采用前文提出式4.8，指定深度 CNN 模型中不同卷积层中不同通道，利用前文提出的带动量的梯度更新策略，可视化结果见图4.4，其中从上到下依次为 GoogleNet 模型低中高层不同通道的可视化结果，与文献^[18]一致，低层多尺度分辨率生成的纹理见图4.4首行所示，中层是一些物体部件，见图 4 中间行所示蜜蜂的局部结构，而高层是更完整的抽象概念见图4.4下层中完整的花瓣。对比图4.4(b)、(c)列，可验证拉普拉斯金字塔主动分解提升图像部分低频成分，而高斯金字塔分解生成的图像中高频细节更突出。

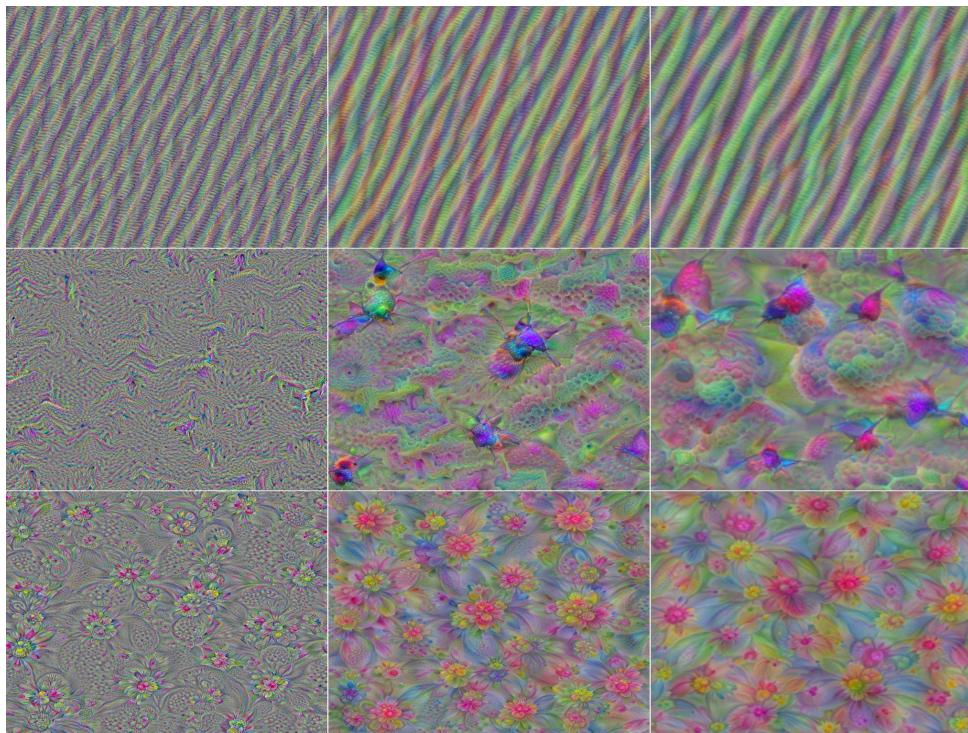


图 4.4: 金字塔分解正则化可视化效果

4.5.2 引入类别显著性的可视化

通过观察之前可视化结果可知，生成的图像中除了该类别外仍有许多额外的上下文信息（见图4.2中鸟类别的树枝），这些信息与模型的分类能力相关联，可通过引入类别激活图可改善可视化效果。迭代更新过程中依据采用式4.11，使用类别激活图作为加权因子限制迭代更新区域。

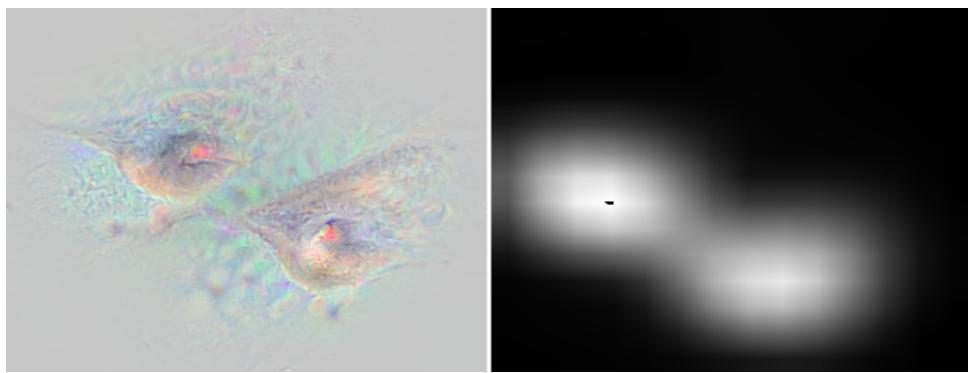


图 4.5: 引入类别激活图的可视化

实验结果见图4.5(a) 所示，具体实验设置和图 2 采用的参数一致，使用提出的金字塔分解正则化技术，图4.5(b) 为图4.5(a) 相应的类别激活图，图4.5(a) 结果表明与类别无关的上下文信息得到压制，但仍存在两个类别中心。

4.6 本章小结

本文针对理解深度 CNN 特征空间存在的问题，提出一种用于改善深度 CNN 分类模型的可视化方法。其中通过改善激活最大化可视化技术来产生更具有全局结构的细节、上下文信息和更自然的颜色分布的高质量图像。该方法首先对反向传播的梯度进行归一化操作，在常用正则化技术的基础上，提出使用空间金字塔分解图像不同频谱信息；为限制可视化区域，提出利用类别显著激活图技术，可以减少优化产生重复对象碎片的倾向，而倾向于产生单个中心对象以改进可视化效果。激活最大化可显示 CNN 在分类时关注什么。这种改进的深度可视化技术将增加我们对深层神经网络的理解，进一步提高创造更强大的深度学习算法的能力。该方法适用于基于梯度更新的可视化领域，是对网络模型整体的理解，具体各层特征怎么耦合成语义信息仍需进一步探索，深度 CNN 模型如何重建一个完整的类别概念，仍是一个开放性问题。

第五章 医学计算机辅助检测方法

提出了一种基于深度卷积神经网络自动识别超声心动图标准切面的方法，并可视化分析了深度模型的有效性。该算法针对网络全连接层占有模型大部分参数的缺点，引入空间金字塔均值池化替代全连接层，获得更多的空间结构信息，并大大减少模型参数、降低过拟合风险，通过类别显著性区域将类似注意力机制引入模型可视化过程。通过超声心动图标准切面的识别问题案例，试着对深度卷积神经网络模型的鲁棒性和有效性进行了解释。在超声心动图上的可视化分析实验表明，通过改进方法的深度模型的识别决策依据，同医师辨别分类超声心动图标准切面的依据一致，表明了方法的有效性和实用性。

5.1 引言

计算机辅助检测 (Computer-Aided Detection, CADs) 是医学影像诊断过程中的一项重要任务，是进行相关结构功能测量的前提条件。其中，二维图像的目标组织结构自动检测是 CADs 技术的核心基础。在临床实践中，医生需整合不同模态、不同位置方向且以不同比例显示的图像信息，目前的研究主要关注如何使检测过程快速自动化。由于医学影像自身的特殊性，比如缺乏大量高质量标注数据；大多数医学目标组织结构存在非刚性形变；图像背景前景的区分不明显等，导致组织结构自动定位比较困难。现有大多数 CADs 系统在临床实际应用中表现不佳的原因是：检测结果的敏感性和特异性都较低，诊断效能低^[46]。

不同模态的医学图像中，如超声、计算机断层扫描 (Computed Tomography, CT) 和核磁共振 (Magnetic Resonance Imaging, MRI) 等，都存在目标身体器官自动定位的问题。以左心室 (Left Ventricle, LV) 检测为例，大多数 LV 定位方法主要依据位置，时间和形状的假设。基于位置的方法仅假设心室在图像的中心，该方法并不对不同病人心室位置的差异性以及图像的尺寸变化进行考虑，效果较差；基于时间的方法，假设左心室是图像中唯一的运动对象，然而这种方法敏感性高，除心室的运动伪影之外，还存在其它运动的器官，如 Schollhuber^[47] 针对 MRI 短轴使用时空信息并消除运动伪影，由分层模式匹配算法定位包含 LV 的感兴趣区域，其通过使用互信息图像配准使运动伪影最小化，随后估计特征强度—时间曲线进行像素分类和边界的提取，得到最终分割结果；基于形状的方法将 LV 视为圆 (短轴)、椭圆 (长轴)，然而该方法通常针对异常形状的 LV 容错性差，如 Lu 等^[48] 使用大津阈值度量圆形程度，然后进行霍夫变换定位 LV 位置。也可搜索每个切片的质心，并用三维最小二乘拟合去除异常值，得到分割结果^[49]。

不依据具体的强先验假设，机器学习算法可通过区分前景目标对象和背景来

解决目标结构自动检测的问题。如 Kellman 等^[50] 提出了一种使用概率集成提升树来估计 LV 姿态和用空间间隔学习 LV 短轴边界的方法。Zhou 等^[51] 在超声心动图中通过规一化集成提升回归学习非线性映射以定位 LV，其团队后来提出针对多个器官的特异性置信最大化分类器，整合更高的自由度以改善回归定位任务的精度。Liu 等^[52] 通过利用基于子模块函数优化理论的多标记搜索策略来进行标记点的检测。Zheng 等^[53] 在实现器官定位的同时，通过组合优化置信度来估计目标器官的位置、缩放及朝向等参数值。前述机器学习算法都基于弱先验知识，启发式设计相关特征，结合滑动窗口策略，选择分类器进行分类判断窗口中内容以估计相应位置。

近来通用物体检测领域取得巨大进展，主要得益于深度学习能利用大量标注数据，从原始像素出发，逐层分级学习中高层抽象语义特征^[12]。区域卷积神经网络^[54] 在大规模自然图像数据集（如 ImageNet^[11]）上，识别性能远超传统方法^[2,54]。当前实践中由于深度学习需要大量的训练数据，所以仅在少数医学任务中取得有限的成功应用。深度学习方法用在定位检测问题时可分为两个阶段^[55]：候选框位置选取和窗口内容类别分类。如利用深度卷积网络进行显微镜图像中细胞检测^[56]、结合深度全卷积网络的 MRI 心室检测与分割^[57,58] 和超声图像解剖结构的检测^[59]。这些方法大都关注特定目标结构的检测分割，而本文专门针对目前 CADs 普遍存在的检测定位问题，基于改进的生成候选框的快速区域深度卷积神经网络（Faster RCNN）^[60] 方法，提出一种医学目标结构检测框架：1) 在区域生成网络的基础上引入空间变换损失使得候选框生成网络能捕捉目标的空间变换参数；2) 采用在线困难样例挖掘策略，加快训练收敛过程，提高检测小目标的准确度；3) 并基于目标先验知识，针对左心室提出利用检测二尖瓣环、心内膜垫和心尖位置，高效估计左心室姿态参数。4) 为验证该算法的鲁棒性和有效性，分别针对两个具体 CADs 应用进行实验分析。

5.2 区域卷积神经网络概览

5.2.1 物体检测形式化定义

若用 r 来表示图像中的矩形窗口区域，令 R 表示由对象检测系统提供的所有候选窗口的集合，将有效定位标记定义为 R 的子集，使得标记位置内内容“不重叠”，令 Y 来表示所有有效标记位置的集合。并合并常用的非最大值抑制（Non-maximum suppression, NMS）过程，给定图像 x 和窗口评分函数 f ，物体检测算法流程可定义为：

形式化定义物体检测过程见公式 5.1，式中参数定义请参考算法 5.1。

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.1)$$

表 5.1: 算法 1 物体检测

Input: 图像 x , 窗口得分函数 f
1: $D :=$ 所有候选框 $r \in$ 使得 $f(x, r) > 0$
2: 按 f 排序 D 使得 $D_1 \geq D_2 \geq D_3 \geq \dots \geq D_n$
3: $y^* :=$
4: for $i = 1$ to n do
5: 若 D_i 和 y^* 中任意候选框不重叠
6: $y^* := y^* \cup D_i$
7: end for
8: Return: y^* , 物体的目标位置.

通常公式5.1可通过贪心搜索的方法来完成，算法将联合最小化在算法5.1中产生假阳例的数量和最大化检测窗口评分函数，即寻找具有最大得分但同时不重叠的滑动窗口位置集合。

5.2.2 区域卷积神经网络的演进

2014 年 Girshick 等^[54] 提出区域卷积神经网络 (Region-based Convolutional Neural Network, RCNN)，对每一候选框窗口都进行一次前向传播，这将导致冗余计算，时间复杂度高，为解决这一问题，He 和 Ren 等提出 SPP-net^[?] 和 Fast RCNN^[55] 加以改进，不再把每一候选窗口均送入网络，而是仅对图像特征提取一次，把原图中候选区域投影到卷积特征图上，然后对投影后的区域特征图进行空间感兴趣区域池化 (ROI Pooling) 得到固定长度的特征向量。其中 Fast RCNN 中的兴趣区域池化是 SPP-Net 中多尺度空间金字塔池化的特例，仅用单一尺度的金字塔池化操作。RCNN 及其改进的 Fast RCNN 都依赖于人为设计的候选框生成方法，如选择性搜索等。为减少生成候选框的计算时间，Faster RCNN 提出区域生成网络 (Region Proposal Networks, RPN)，区域生成网络和检测网络共享提取特征的卷积层，仅提取几百个或者更少的高质量预选窗口，且召回率较高（导致更少的假阳例）。但现有的通用物体检测算法均是假设候选框为矩形，不能解决旋转朝向问题。

5.3 候选区域生成网络及其改进

本章将分别从候选区域生成网络模型的结构、仿射变换候选框区域的生成、空间变换损失函数的设计、模型训练方法等方面介绍本文所提出框架，并结合 Faster R-CNN 模型提出端到端的目标检测方法。

5.3.1 候选区域生成网络模型结构

候选区域生成网络将一图像（任意大小）作为输入，输出目标候选框的集合和每个候选框内有无目标的概率估计，如图5.1右图所示，RPN 在卷积层后接两个全卷积层完成候选区域生成功能，以实现增加滑动窗口操作。该模型使用全卷积网络 [20] 处理任意大小的图片输入，为了和目标检测网络^[55] 共享计算，在特征提取的过程中同时计算目标检测所需的感兴趣区域的初始估计，在最后一个共享卷积层输出的特征映射图上滑动小网络，卷积特征映射图上 $n \times n$ 大小空间窗口作为该网络全连接的输入，本文 n 取 3。每个滑动窗口映射到一个低维向量上（如图一左上中 256-d），该向量输出给两个全连接层——候选框位置回归层和候选框类别分类层。原文中采用类别无关分类损失，即仅区分该候选框内是否包含物体（前/背景），本文将其扩展为类别相关的分类损失。

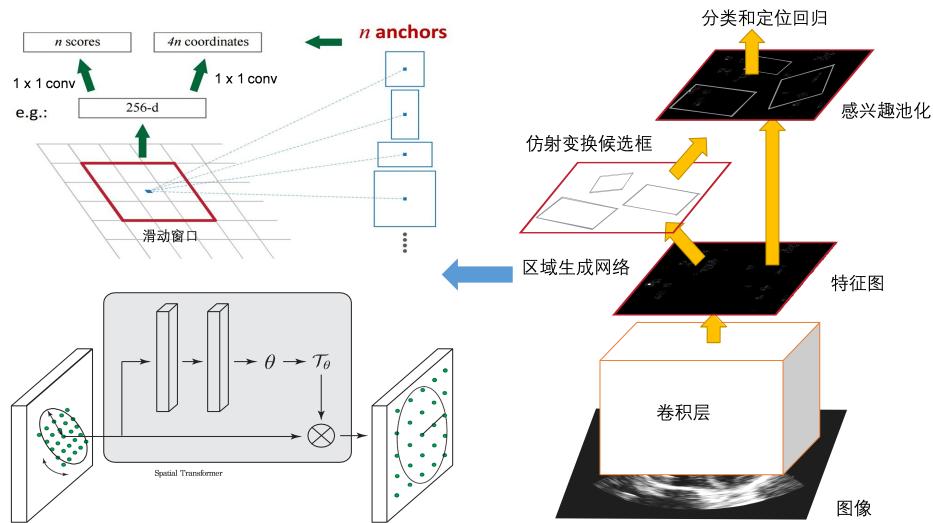


图 5.1: 左上: 引入空间不变性的 anchor 机制 左下: 空间变换网络右: Faster R-CNN 带仿射变换的检测模型框架

为引入空间尺度不变性，采用多尺度和多纵横比的“参考”框(anchor)(图5.1左上所示)。该机制可看作是金字塔型参考框的回归，避免了枚举多尺度、多纵横比的图像或卷积核。在每一个滑动窗口的位置，同时预测 k 个参考区域，回归层有 $4k$ 个输出，即 k 个 box 的坐标编码，多元逻辑回归分类层输出 $(c+1) \times k$ 个(物体

类别数 c 加背景类的) 概率估计。候选框由相应的 k 个 anchor 的参数化表示, 每个 anchor 以当前滑动窗口中心为中心, 并对应一种尺度和长宽比, 我们使用 3 种尺度和 3 种长宽比, 在每一个滑动位置就有 $k=9$ 个 anchor。对于大小为 $w \times h$ 的卷积特征映射, 总共有 $w \times h \times k$ 个 anchor。

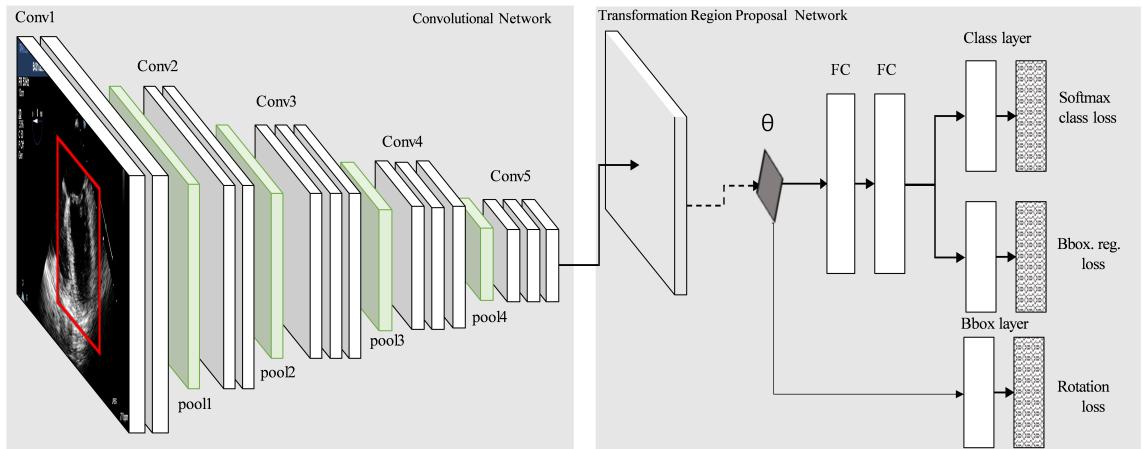


图 5.2: 考虑物体朝向的区域生成网络模型结构示意图, 图中 conv 表示卷积层, pool 表示池化层, 右侧的带仿射变换的区域生成网络模型同最终的检测模型一致, 其中的 FC 表示全连接层, softmax class loss 表示多任务损失中的分类损失, Bbox.reg loss 表示候选框回归定位损失, Rotation loss 表示文中的针对变换参数 θ 的 Von Mise 损失。

5.3.2 仿射变换候选框

为检测物体的姿态, 结合空间变换网络^[61] (见图5.1左下), 提出带仿射变换的候选框生成算法。之前的候选框生成方法仅考虑固定尺度和宽高比的矩形框, 并未考虑物体的旋转朝向, 二维空间仿射变换可表示为:

$$\begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix} = \tau_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \quad (5.2)$$

式中 $(x_i^s, y_i^s)^T$ 为输入特征图中目标坐标系下的网格点, τ_θ 为变换矩阵, $(x_i^t, y_i^t)^T$ 为输出特征图中目标坐标系下的采样网格点。其中由于图像的坐标不是中心坐标系, 宽高坐标需归一化表示, 如 $-1 \leq x_i^s, y_i^s \leq 1$, 且采用图形学中齐次坐标表示。公

式5.2能用六个参数定义对输入特征图的裁剪、平移、旋转和缩放等变换。该公式进步简化为只考虑旋转变换：

$$\begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix} = \tau_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} \quad (5.3)$$

其中 α 表示绕图像中心顺时针旋转角度，通常变换后的像素并不是在相应网格的整数值，常用双线性插值进行近似，变换后的候选框送入感兴趣区域池化层，后接多任务损失函数。实质是把空间变换层嵌入到 RPN 网络中，并且引入有监督的损失以指导空间变换。

5.3.3 朝向回归损失函数

旋转朝向的周期性会导致两个问题：(1) 要优化的损失函数不能区分对于周期性损失，简单地将模运算符应用于网络的输出会导致不可靠的损失，不能再被鲁棒地优化。(2) 由大多数参数模型中执行的矩阵向量积产生的回归输出是固定的线性运算。为此提出旋转朝向回归损失 $L_{VM}(o, o^*)$ ，第一个问题可以通过采用 Von Mise 分布^[62] 来解决损失函数不连续性，其近似服从于单位圆上的正态分布：

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.4)$$

其中 p 指相应的概率密度函数，指角度，是分布的平均角度，与近似高斯方差成反比，而是阶数为 0 的修正贝塞尔函数，利用余弦函数来避免不连续性，可以得出以下损失函数：

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.5)$$

式中 t 为预测旋转角度大小， t 为真实旋转角度大小，称 t 为目标值， k 为控制损失函数尾部的简单超参数。由角度正余弦组成的二维向量替代表示，利用自然语言处理文献中广泛使用的余弦代价函数 [31] 来解决使用线性操作来预测周期值的问题：

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.6)$$

在神经网络框架中的实现是相对简单的，因为所需要的是全连接层和归一化层，前向传播公式：

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.7)$$

式中 θ 和 ϕ 是来自全连接层的可学习参数，然后反向传播归一化损失的导数为

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.8)$$

式中归一化确保输出值被联合学习，通过比较 CVM 和 Ccos，最终朝向回归损失函数为

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.9)$$

与式 6 相似，主要区别在于存在 e ，它将目标值附近的错误“下推”，实际上是较小地惩罚小错误。

5.3.4 带朝向的多任务损失函数

多任务损失分别存在于 RPN 及检测网络中，图 2 中显示的是检测网络结构示意图。每一个候选框均送感兴趣池化层，后接两层的全连接层和多元逻辑回归分类损失（图 2 中 Softmax loss），候选区域回归定位损失（图 2 中 Box.reg loss）和旋转朝向回归损失（图 2 中 Rotation loss）：式中，分别代表预测类别分类概率，候选框偏移量和感兴趣区域内物体的朝向大小；表示标记类别为背景，表示框内是否有目标的指示函数，分别表示物体的候选框标记和真实朝向。为两个损失的相应平衡权重大小，详细形式如下：

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.10)$$

(11)

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.11)$$

(12)

$$y^* = \arg \min_{y \in Y} \sum_{r \in Y} f(x, r) \quad (5.12)$$

(13) 和是公式 4 中的分类损失和相应的平滑 L1 损失， c 代表类别数。

5.3.5 困难样例挖掘

由于医学数据样本标注困难，数量相对较少，一般假设与目标位置矩形框有重叠的候选框是有较大概率是难以区分的，结果也可能是次优的，因为在其他位置可能存在更难区分的样本，导致模型收敛变慢，误警率高。在每次迭代训练过程中采用在线困难样例挖掘方法（Online Hard Example Mining, OHEM）[23]，对所有候选框的损失进行排序，由于相似候选框重叠区域的损失很接近，可采用非极大值抑制策略限制候选框的数目，选择前 k 个最大损失作为困难样例，反向传播其相应的梯度，其他候选框的梯度不进行回传，即不更新模型权重。

候选区域生成网络模型结构

5.4 实验结果分析和讨论

为验证提出的自动检测算法的有效性和正确性,本节将分别采用一个公开可用的 MRI 数据集,和我们收集的来源于四川大学华西医院麻醉科的经食道超声心动图数据集(不包含患者信息)上进行实验。相关实验代码请参考 <https://github.com/taopanpan/echo>。

3.1 检测 MRI 左心室短轴纽约大学提供的公用数据集 [24] 包含 33 名患者的心脏 MRI 体数据,以及 LV 心内膜和心外膜的手动分割结果。该数据集中的大多数切片包含心脏疾的病例切片。该数据集使用 GE Genesis Signa MRI 扫描仪,采取 FIESTA 方案扫描获得。每个患者的 20 个序列帧包含 8-15 个短轴切片,大小为 256x256,厚度为 6-13 mm,像素分辨率为 0.93-1.64 mm。为了检验所提出方法定位性能,取 14 个体数据形成 1176 个切片作为训练集,其余作为测试集。本实验中不使用旋转朝向损失,评价指标采用文献 [15] 中定量评估计算左心室短轴 (SAX) 定位的准确度,敏感性和特异性。表 1 不同模型检测精度的比较

为评价不同深度模型对检测效果的影响,实验的检测模型选取 VGG16[25] 和 ResNet101[26],训练方法采取端到端的近似联合优化,OHEM 表明训练过程中采用困难样例挖掘方法,即在训练中只选择损失占前 70 另一方面,敏感性是最容易提高的参数,平均超过 8%,模型不能正确定位为大尺寸的心脏,导致较小 LV 切片的高 FP,降低了整体系统性能。而困难样例挖掘的方法没有显著提高特异性,因为 TN 和 FP 都降低。考虑到心脏异常的高变异性导致心脏形状的大变异性,所提出的算法均能成功定位 LV 短轴,当检测出心室短轴时,可大致确定心室中心点(如图 5(a) 所示),利用二腔心 (2CH) 和四腔心切面 (4CH) 均垂直于短轴切面的先验,找到与 SAX 的 2CH 和 4CH 交集在 SAX 平面上投影,然后得到投影线在 2D 图像上相交的位置,即为左心室的 3D 位置(如图 5(b) 所示)。3.2 检测左心室及其朝向 MRI 左心室短轴的检测由于组织结构相对简单,且噪声少。为验证提出算法的通用性,针对超声图像左心室长轴切面检测心室、二尖瓣环、心内膜垫和心尖位置,并估计左心室朝向。主要包含单扇形和多普勒成像的双扇形两种由专业医师标注食管中段四腔心 (ME4C) 的标准切面视频构成,视频中至少包含 2-3 个心动周期,依据医师建议从视频中截取 5 帧,并经医师检验手工筛选后得到 900 张 ME4C 切面,对切面内左心室 (LV),二尖瓣环、心内膜垫和心尖位置进行人工标注作为“金标准”。其中随机选取 100 张作为测试集,其余作为训练集。训练时采用提出的联合多任务损失,以 VGG16 网络作为检测的预训练的模型为例,在 RPN 中添加空间变换网络实现了各个候选框的空间变换,并施加旋转朝向损失。VGG16 网络特征提取器包括 13 个卷积层,并输出 512 个 conv5 特征图,空间变换网络包括具有两个同样卷积池化层组成的定位网络,其由 20 个卷积核大小为 5、步长为 1 和核大小为 2 的池化层构成,两层全连接层回归得出 6 个仿射变换参数,其中,全连接层的激活函数需选择为双曲正切函数,权重高斯初始化,而变换参数

初始化为 $[1 \ 0 \ 0 \ 0 \ 1 \ 0]^T$ 。其它跟 Faster RCNN 中设置一致，其中，分别取 0.1 和 0.001；训练方法采取端到端的近似联合优化，迭代轮数为 50000。评价指标采用平均检测精度 (mean average precision, mAP)，是多个类别平均检测精度的平均值。表二显示使用提出方法分别在 VGG16 模型和 ResNet101 模型上，结合困难样例挖掘训练方法得出的测试结果，其中 OHEM 表示相应模型结合在线困难样例挖掘方法的检测结果，STN 表示结合提出带朝向损失的空间变换网络的检测结果，在测试集上，针对左心室的 AP 最优可达 99.1 表 2 不同模型检测精度，LV 表示左心室，Apx 代表心尖，left 代表二尖瓣环，right 代表心内膜垫 Tab.2 Different models of detection accuracy, LV on behalf of the left ventricle, Apx on behalf of the apex, left said mitral annulus, right on behalf of the endocardial cushion method mAP LV apx left right

为验证提出算法在检测左心室位置的同时可以回归学习左心室的姿态参数、预测左心室的朝向变换，超参数 k 跟文献 [22] 一致，交叠比大于 0.5 时估计姿态参数，人为标定心室朝向存在较大偏差，但可以根据二尖瓣环、心内膜垫和心尖位置估算出心室朝向角度作为对照。由于 ME4C 切面中心室的大概朝向的分布范围在 $[-45^\circ, 45^\circ]$ 之间，通过手工构建训练集，训练样本旋转以 15° 为间隔的指定角度。通过分析相关估算结果和预测结果，可以发现二者具有很大的一致性。左心室检测结果和旋转朝向结果见表 3，检测结果如图 5(c,d) 所示，更多实验结果请参考给定开源地址。表 3 不同旋转角度分类检测性能比较，Compute 表示根据额外标记计算得到的结果，Pred 表示模型预测结果 Tab.3 Comparison of different rotation angle detection performance, Compute represents the results calculated according to the additional markers, Pred represents the model prediction results method -45o -30o -15o +15o +30o +45o Avg Compute 66.6 78.0 87.3 85.5 75.8 62.3 75.9 Pred 73.0 81.7 81.7 89..5 80.4 70.3 80.7

为了更详细地评估模型性能，使用检测分析工具 [27] 分析了心尖位置的检测结果，如图 3 显示模型可以准确（白色区域）检测到心尖位置，召回率在 84-87% 左右，并且比“弱”标准（小于 0.1 交叠比）高得多。针对心尖位置的定位精确度较低，这是因为医师在标定心尖位置时有很大的随意性，且目标尺寸较小，与类似对象类别有更多的混淆。(a) (b) 图 3 (a) 显示 apx 检测精度的累积分布：正确的 (Cor) 或定位不准确 (Loc) 的假阳性，与之混淆类似类别 (Sim) 与其他类别 (Oth) 或背景 (BG)。固体红色线是以“强”标准（大于 0.5 交叠比），反映精确度随检测增加而变化。红色虚线使用“弱”标准（大于 0.1 交叠比）。(b) 显示排名靠前的假阳性类型的分布。Fig. 3: Visualization of performance for our model on apx. (a) shows the cumulative fraction of detections that are correct (Cor) or false positive due to poor localization (Loc), confusion with similar categories (Sim), with others (Oth), or with background (BG). The solid red line reflects the change of recall with

” strong” criteria (0.5 overlap) as the number of detections increases. The dashed red line is using the ” weak” criteria (0.1 overlap). (b) shows the distribution of top-ranked false positive types.

(a) (b) (c) (d) 图 5 (a,b) 表示不同 MRI 图像检测左心室结果, (c,d) 两图表示超声心动图的 ME4C 切面的左心室、二尖瓣环、心内膜垫和心尖位置及旋转角度的检测结果。Fig. 5 (a,b) show the results of the left ventricular of different MRI images. (c,d) show the left ventricular, mitral annulus, endocardial pad and apical position and rotation angle of the ME4C section of echocardiography of the test results.

5.5 本章小结

4 结语本文利用深度学习来解决医学图像计算机辅助检测问题, 设计并验证了自动检测 MRI 短轴和超声心动图中 LV 长轴切面的方法, 在通用物体检测 Faster RCNN 框架的基础上, 针对 RPN 引入空间变换, 结合带朝向损失的多任务损失, 探索解决图像平面内物体旋转角度检测的问题, 并利用困难样例挖掘策略加快迭代训练。在公共 MRI 数据集和自主收集的超声心动图数据上进行详尽实验验证, 在多个评估指标方面提供更好的测试结果, 但该方法仍耗费较多的标注数据, 探索需要更少标注数据的检测算法是将来的工作目标。

第六章 医学图像的分割方法

提出了一种基于深度卷积神经网络自动识别超声心动图标准切面的方法，并可视化分析了深度模型的有效性。该算法针对网络全连接层占有模型大部分参数的缺点，引入空间金字塔均值池化替代全连接层，获得更多的空间结构信息，并大大减少模型参数、降低过拟合风险，通过类别显著性区域将类似注意力机制引入模型可视化过程。通过超声心动图标准切面的识别问题案例，试着对深度卷积神经网络模型的鲁棒性和有效性进行了解释。在超声心动图上的可视化分析实验表明，通过改进方法的深度模型的识别决策依据，同医师辨别分类超声心动图标准切面的依据一致，表明了方法的有效性和实用性。

6.1 引言

6.2 初始位置定位和特征点标注

6.3 AAM 模型和 CLM 模型

6.4 结合卷积网络特征的形状对齐模型

6.4.1 超声组织特征纹理特异性灰度归一化

6.4.2 结合不同外观特征的全局 AAM

6.5 实验结果分析和讨论

6.6 本章小结

本文利用深度学习来解决医学图像计算机辅助检测问题，设计并验证了自动检测 MRI 短轴和超声心动图中 LV 长轴切面的方法，在通用物体检测 Faster RCNN 框架的基础上，针对 RPN 引入空间变换，结合带朝向损失的多任务损失，探索解决图像平面内物体旋转角度检测的问题，并利用困难样例挖掘策略加快迭代训练。在公共 MRI 数据集和自主收集的超声心动图数据上进行详尽实验验证，在多个评估指标方面提供更好的测试结果，但该方法仍耗费较多的标注数据，探索需要更少标注数据的检测算法是将来的工作目标。

第七章 总结与展望

参考文献

- [1] JARRETT K, KAVUKCUOGLU K, RANZATO M, et al. What is the best multi-stage architecture for object recognition?[C/OL]//Proceedings of the IEEE International Conference on Computer Vision. 2009: 2146–2153.
- [2] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[C]//Advances In Neural Information Processing Systems. 2012: 1–9.
- [3] CHEN H, NI D, QIN J, et al. Standard Plane Localization in Fetal Ultrasound via Domain Transferred Deep Neural Networks[J]. IEEE Journal of Biomedical and Health Informatics, 2015, 19(5): 1627–1636.
- [4] EBADOLLAHI S, CHANG S F C S F, WU H. Automatic view recognition in echocardiogram videos using parts-based representation[J]. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., 2004, 2.
- [5] Kevin Zhou S, PARK J H, GEORGESCU B, et al. Image-based multiclass boosting and echocardiographic view classification[C/OL]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition: volume 2. 2006: 1559–1565.
- [6] OTEY M, BI J, KRISHNA S, et al. Automatic view recognition for cardiac ultrasound images[C/OL]//MICCAI: International Workshop on Computer Vision for Intravascular and Intracardiac Imaging. 2006: 187–194.
- [7] ROY A, SURAL S, MUKHERJEE J, et al. Modeling of Echocardiogram Video Based on Views and States[M/OL]//Computer Vision, Graphics and Image Processing. Springer Berlin Heidelberg, 2006: 397–408.
- [8] PARK J H, ZHOU S K, SIMOPOULOS C, et al. Automatic Cardiac View Classification of Echocardiogram[C/OL]//2007 IEEE 11th International Conference on Computer Vision. IEEE, 2007: 1–8.
- [9] BEYMER D, SYEDA-MAHMOOD T, WANG F. Exploiting spatio-temporal information for view recognition in cardiac echo videos[C/OL]//2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2008: 1–8.
- [10] WU H, BOWERS D M, HUYNH T T, et al. Echocardiogram view classification using low-level features[C]//Proceedings - International Symposium on Biomedical Imaging. 2013: 752–755.
- [11] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211–252.

- [12] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: An astounding baseline for recognition[C/OL]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2014: 512–519.
- [13] BAR Y, DIAMANT I, WOLF L, et al. Chest pathology detection using deep learning with non-medical training[C]//IEEE International Symposium on Biomedical Imaging. [S.l.: s.n.], 2015: 294–297.
- [14] OLIVA A, TORRALBA A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. International Journal of Computer Vision, 2001, 42(3): 145–175.
- [15] MARGETA J, CRIMINISI A, Cabrera Lozoya R, et al. Fine-tuned convolutional neural nets for cardiac MRI acquisition plane recognition[J/OL]. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 2017, 5(5): 339–349.
- [16] SIMONYAN K, VEDALDI A, ZISSERMAN A. Deep Inside Convolutional Networks Visualising Image Classification Models and Saliency Maps[C/OL]//Int. Conf. Learn. Represent. 2014: 1–8.
- [17] MAHENDRAN A, VEDALDI A. Understanding deep image representations by inverting them[C/OL]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015: 5188–5196.
- [18] ZEILER M D, FERGUS R. Visualizing and Understanding Convolutional Networks [M/OL]//FLEET D, PAJDLA T, SCHIELE B, et al. Computer Vision – ECCV 2014: 13th European Conference: 8689 LNCS. Zurich: Springer International Publishing, 2014: 818–833.
- [19] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning Deep Features for Discriminative Localization[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition Learning: volume 111. 2015: 2921–2929.
- [20] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[J/OL]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, 07-12-June-2015(2): 1–9.
- [21] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C/OL]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): volume 7. 2016: 770–778.
- [22] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916.
- [23] CHATFIELD K, SIMONYAN K, VEDALDI A, et al. Return of the Devil in the Details: Delving Deep into Convolutional Nets[C/OL]//Proceedings of the British

- Machine Vision Conference 2014. British Machine Vision Association, 2014: 6.1–6.12.
- [24] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[J/OL]. Proceedings of the ACM International Conference on Multimedia, 2014: 675–678.
 - [25] ERHAN D, BENGIO Y, COURVILLE A, et al. Visualizing higher-layer features of a deep network[R/OL]//Univ. Montr. Montréal, Canada, 2009: 1–13.
 - [26] LENCI K, VEDALDI A. Understanding image representations by measuring their equivariance and equivalence[C/OL]//2015 IEEE Conf. Comput. Vis. Pattern Recognit. IEEE, 2015: 991–999.
 - [27] SZEGEDY C, ZAREMBA W, SUTSKEVER I. Intriguing properties of neural networks[C/OL]//Int. Conf. Learn. Represent. 2014: 1–10.
 - [28] YOSINSKI J, CLUNE J, NGUYEN A, et al. Understanding Neural Networks Through Deep Visualization[C/OL]//Deep Learning Workshop, International Conference on Machine Learning (ICML). 2015.
 - [29] NGUYEN A, DOSOVITSKIY A, YOSINSKI J, et al. Synthesizing the preferred inputs for neurons in neural networks via deep generator networks[J/OL]. arXiv, 2016: 1–29.
 - [30] NGUYEN A, YOSINSKI J, CLUNE J. Multifaceted Feature Visualization: Uncovering the Different Types of Features Learned By Each Neuron in Deep Neural Networks [C/OL]//Proc. Work. Vis. Deep Learn. Int. Conf. Mach. Learn. 2016: 23.
 - [31] DOSOVITSKIY A, BROX T. Inverting Visual Representations with Convolutional Networks[C/OL]//2015 IEEE Conf. Comput. Vis. Pattern Recognit. 2015: 184–199.
 - [32] MAHENDRAN A, VEDALDI A. Visualizing Deep Convolutional Neural Networks Using Natural Pre-images[J/OL]. Int. J. Comput. Vis., 2016: 1–23.
 - [33] CAO C, LIU X, YANG Y, et al. Look and Think Twice: Capturing Top-Down Visual Attention with Feedback Convolutional Neural Networks[C/OL]//2015 IEEE Int. Conf. Comput. Vis. IEEE, 2015: 2956–2964.
 - [34] BACH S, BINDER A, MONTAVON G, et al. Analyzing Classifiers: Fisher Vectors and Deep Neural Networks[C/OL]//Proc. IEEE Conf. Comput. Vis. Pattern Recognit. 2016: 17.
 - [35] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and Harnessing Adversarial Examples[C/OL]//Int. Conf. Learn. Represent. San Diego, USA, 2014: 484–485.
 - [36] HUANG B. FaceNet : A Unified Embedding for Face Recognition and Clustering [C]//2015 IEEE Conf. Comput. Vis. Pattern Recognit. Boston, USA: [s.n.], 2015: 815–823.
 - [37] DENTON E, CHINTALA S, SZLAM A, et al. Deep Generative Image Models using a

- Laplacian Pyramid of Adversarial Networks[C/OL]//Adv. Neural Inf. Process. Syst. 28. 2015: 1486—1494.
- [38] KRISHNAN D, FERGUS R. Fast image deconvolution using hyper-laplacian priors [C/OL]//Y. Bengio, SCHUURMANS D, LAFFERTY J D, et al. Adv. Neural Inf. Process. Syst.: volume 28. Curran Associates, Inc., 2009: 1–9.
- [39] BURT P, ADELSON E. The Laplacian Pyramid as a Compact Image Code[J]. IEEE Trans. Commun., 1983, 31(4): 532–540.
- [40] VAN DER SCHAAF A, VAN HATEREN J H. [M].
- [41] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C/OL]//Proc. 32nd Int. Conf. Mach. Learn.: volume 37. Lille, France, 2015: 448—456.
- [42] DUCHI J, HAZAN E, SINGER Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization[J/OL]. J. Mach. Learn. Res., 2011, 12: 2121–2159.
- [43] GATYS L A, ECKER A S, BETHGE M. Texture Synthesis Using Convolutional Neural Networks[J]. Adv. Neural Inf. Process. Syst., 2015: 262–270.
- [44] HE K, WANG Y, HOPCROFT J. A Powerful Generative Model Using Random Weights for the Deep Image Representation[C/OL]//Adv. Neural Inf. Process. Syst. 28. Barcelona, Spain, 2016: 1–8.
- [45] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C/OL]//Int. Conf. Learn. Represent. San Diego, USA, 2015: 1–14.
- [46] CHENG J Z, NI D, CHOU Y H, et al. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans[J/OL]. Sci. Rep., 2016, 6(1): 24454.
- [47] SCHÖLLHUBER A. Fully Automatic Segmentation of the Myocardium in Cardiac Perfusion MRI[J/OL]. Engineering in Medicine, 2008, 3(22): 12–19.
- [48] LU Y, RADAU P, CONNELLY K, et al. Segmentation of Left Ventricle in Cardiac Cine MRI: An Automatic Image-Driven Method[M/OL]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009: 339–347.
- [49] PETITJEAN C, DACHER J N. A review of segmentation methods in short axis cardiac MR images[J/OL]. Med. Image Anal., 2011, 15(2): 169–184.
- [50] KELLMAN P, LU X, JOLLY M P, et al. Automatic LV localization and view planning for cardiac MRI acquisition[J/OL]. J. Cardiovasc. Magn. Reson., 2011, 13(Suppl 1): P39.
- [51] ZHOU S K, GEORGESCU B, ZHOU X S, et al. Image based regression using boosting method[C/OL]//Proc. IEEE Int. Conf. Comput. Vis.: I. 2005: 541–548.
- [52] SHE Y, LIU D C. An Interactive Editing Tool from Arbitrary Slices in 3D Ultrasound Volume Data[M/OL]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007: 2447–2451.

- [53] ZHENG Y, COMANICIU D. Marginal Space Learning for Medical Image Analysis [M/OL]. New York, NY: Springer New York, 2014: 199–256.
- [54] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C/OL]//2014 IEEE Conf. Comput. Vis. Pattern Recognit. IEEE, 2014: 580–587.
- [55] GIRSHICK R. Fast R-CNN[C/OL]//2015 IEEE Int. Conf. Comput. Vis.: 2015 Inter. IEEE, 2015: 1440–1448.
- [56] AKRAM S U, KANNALA J, EKLUND L, et al. Cell Segmentation Proposal Network for Microscopy Image Analysis[M/OL]//Deep Learn. Data Labeling Med. Appl. 2016: 21–29.
- [57] EMAD O, YASSINE I A, FAHMY A S. Automatic localization of the left ventricle in cardiac MRI images using deep learning[C/OL]//2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE, 2015: 683–686.
- [58] PARK S R, LEE J. A Fully Convolutional Neural Network for Cardiac Segmentation in Short-Axis MRI[J/OL]. arXiv Prepr., 2016: 1–21.
- [59] CHEN H, ZHENG Y, PARK J H, et al. Iterative multi-domain regularized deep learning for anatomical structure detection and segmentation from ultrasound images [J]. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2016, 9901 LNCS: 487–495.
- [60] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[C/OL]//Adv. Neural Inf. Process. Syst. 2015: 1–10.
- [61] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial Transformer Networks[C/OL]//Proc. 28th Int. Conf. Neural Inf. Process. Syst.: volume 25. Montreal, Canada, 2015: 2017–2025.
- [62] BEYER L, HERMANS A, LEIBE B. Biternion nets: Continuous head pose regression from discrete training labels[J]. Lect. Notes Comput. Sci., 2015, 9358: 157–168.

攻读学位期间发表的学术论文与科研成果

已发表论文

1. Pan Tao,Zhongliang Fu,Lili Wang,Kai Zhu.Perceptual Loss with Fully Convolutional for Image Residual Denoising. *Pattern Recognition,CCPR(EI)*. 2016. 122–132,DOI:10.1007/978-981-10-3005-5-11
2. 陶攀, 付忠良, 朱锴, 王莉莉. 金字塔分解的深度可视化方法, 哈尔滨工业大学学报 (EI), 2017,49(11):60-65,DOI:10.11918/j.issn.0367-6234.201612087
3. 陶攀, 付忠良, 朱锴. 基于深度学习的医学计算机辅助检测算法, 生物医学工程学杂志 (EI), 已录用, 2017
4. 陶攀, 付忠良, 朱锴, 王莉莉. 基于深度学习的超声心动图切面识别方法研究, 计算机应用 (中文核心), 2017.DOI:10.11772/j.issn.1001-9081.2017.05.1434
5. Xianghu Ji,Lili Wang,Pan Tao,Zhongliang Fu.Landmark Selecting on 2D Shapes for Constructing Point Distribution Model., *Pattern Recognition,CCPR(EI)* 2016, 318–331.DOI:10.1007/978-981-10-3002-4-27
6. Lili Wang, Zhongliang Fu,Pan Tao.Four-chamber plane detection in cardiac ultrasound images based on improved imbalanced AdaBoost algorithm , *IEEE,ICCCBDA(EI)* 2016,299-303.DOI:10.1109/ICCCBDA.2016.7529574

国家发明专利

1. 纪祥虎, 高思聪, 陶攀, 王莉莉. 用于统计形状模型的特征点辅助标注方法. (申请号:201510672503.8) 专利公开号:CN105205827 A.2015

项目经历

1. 2015-2016 四川省科技创新苗子工程——基于自动分割技术的左心室可视化及功能评价临床教学平台 (编号: 2015060)

项目描述: 本项目主要目标在于使用机器学习方法对左心室进行分割, 得到左心室轮廓及结构和心功能参数; 使用可视化技术对心脏左室进行三维立体结构教学。帮助麻醉医学学员快速学习掌握超声心动图中左心室结构

项目职责: 在项目中主要负责超声图像中心脏器官的自动定位和分割, 分别

利用机器学习的方法对超声图像中的左心室定位，和 AAM 方法对肾脏进行分割。

项目成果：形成论文两篇，专利一项，期间主要研究了基于深度学习的图像预处理方法，基于形状对齐模型进行心室分割，及基于深度级联回归模型进行心室边界分割算法等

2. 2015-2015 阿里巴巴大规模图像搜索赛（38 名共 843 支参赛队伍）

本项目目标是从海量图像中检索最相同或似的 Top20 图像

主要负责使用深度学习模型对图像进行特征抽取，同时配合队友进行图像检索等其他工作，其中用时一个月根据 matconvnet 写了一个 C++ 版本的 CNN 框架的 API，从中获得了处理百万级数据的经验，获得了使用 OpenBLAS 处理大型矩阵运算的经验

项目收获：形成论文一篇，熟悉了深度学习提取语义特征进行实例检索的各项关键技术

3. 2015-2017 四川科技支撑计划-医学图像挖掘与心脏智能诊疗系统关键技术研究

项目描述：本项目主要目标在于使用机器学习方法对超声心动图标准切面进行自动识别。超声图像标准切面分类模块，包括图像预处理、特征提取和分类器模型构建实现标准切面自动识别分类；基于云端的海量切面视频的语义检索模块等

项目职责：项目参与人

任务分工：图像预处理、特征提取、分类器建模、视频语义检索

项目成果：发表论文三篇，分别研究了基于深度学习理论可视化分析其有效性，基于深度特征的超声图像标准切面自动识别算法等

4. 2013-2014 四川省科技支撑项目，华西医院合作项目-医学可视化模拟教学和诊断系统

项目描述：项目旨在为无经验的心脏外科医生和学员提供可视化的教学方案，同时通过机器学习和图形图像处理方法对三维心脏进行开放式建模，以提出一种基于心脏开放模型的智能诊疗综合系统

项目职责：在项目中负责超声图像处理和基于机器学习的病理挖掘工作。

任务分工：图像预处理

项目成果：参与撰写专利两项，对超声仪器，心脏疾病临床基本知识有较全面的了解；设计了针对心脏超声图像的分割识别方法，以及病理挖掘方法；学习了基于偏微分方程的图像去噪和基于水平集的分割方法

在审和 Working 论文

1. 陶攀, 付忠良. 基于 Fast-rcnn 的医学实例检索方法研究, Working, 2015
2. 陶攀, 付忠良. 基于超声心动图的左心室分割综述, Working, 2015
3. 陶攀, 付忠良. 基于形状对齐的超声心动图左心室分割方法, Working, 2016
4. 陶攀, 付忠良. 基于 CNN-LSTM 的超声心动图左心室分割方法, Working, 2017

参与项目编写和申请

1. 2016 四川科技支撑计划—医学图像挖掘与心脏智能诊疗系统关键技术研究
2. 2016 基于医学图像建模的心功能评价系统研发与应用
3. 2015 国科控股技术创新项目—交互式视觉仿真关键技术研究与产品应用示范
4. 2014 西部之光项目—基于医学图像建模的评价系统
5. 2014 数字化医疗辅助设备关键技术研发—基于机器智能的三维可视化手术诊疗仿真平台

获奖及荣誉

1. 2015 获得中国科学院研究生院“三好学生”荣誉称号
2. 2016 中国科学院大学优秀学生干部
3. 2017 中科院博士国家奖学金

致 谢

值此论文完成之际，谨在此向多年来给予我关心和帮助的老师、学长、同学、朋友和家人表示衷心的感谢！

没有 ctex package 的众多前辈的辛勤付出和 CASthesis package 作者吴凌云学长的贡献， \LaTeX 菜鸟的我是无法完成此学位论文模板的。在 \LaTeX 中的一点一滴的成长源于开源社区的众多资料和教程，在此对所有前辈们的付出表示感谢！

.....

谨把本文献给我最敬爱的父亲！